IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Gregory T. Bleck, et al.
Serial No.: 10/759,315 Group No.: 1633
Filed: 1/16/04 Examiner: POPA

Entitled: **PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION**

FIFTH DECLARATION OF DR. GREGORY BLECK

I, Dr. Gregory Bleck, state as follows:

1. My present position is Senior Director, Cell Line Engineering, Catalent Pharma Solutions.

2. I am an inventor of the above referenced patent application.

3. At page 7 of the Office Action, the Examiner recognizes that Mathor and Burns do not teach serial transduction to obtain cells comprising genomes with 20 to 100 integrated vectors. The Examiner then goes on to present arguments as why a person of skill in the art would use serial transduction to obtain cells comprising genomes with 20 to 100 integrated retroviral vectors. The Examiner supports this argument by citations to several publications. However, with one exception, the Examiner has made assumptions that are not factually supported by those references. The one exception is that the references do teach that there is a positive correlation between MOI and integration events. Many of the other assumptions made by the Examiner overstate what the references teach, especially in relation to the invention.

4. At page 7, the Examiner states that Mathor et al. do teach that protein expression is directly proportional to integration events (i.e., copy number)(p. 10376, column 1). The Examiner goes on to state that:

> "It would have been obvious to one of skill in the art, at the time the invention was made, to modify the method of Mathor et al. and Burns et al. by serially transducing their cells with high MOIs (such as MOIs of 1,000) to achieve the claimed ranges of integration

> events, with a reasonable expectation of success. The motivation to do so is provided by Mathor et al., who teach the possibility of specifying the level of transgene expression by controlling the integration events (Abstract, p. 10376, column 1).

The Examiner cites the abstract and p. 10376 of Mathor. Mathor et al. state in this section that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell."

5. The Examiner has overstated the commentary and data in Mathor et al.. It is a fact that Mathor et al. does not contain any data or any statement that the level of transgene expression can be controlled by controlling integration events with the range of 20 to 100 integrations. Mathor et al. presents the data on proviral integration and transgene expression on p. 10373 and in Table 1. This data shows increasing transgene expression as the proviral integrations increase from 1 to 8. When the number of proviral integrations increases to 15, the transgene expression is actually decreases to a level lower than was observed with 8 integrations. Thus, as a factual matter, Mathor et al. teaches that transgene expression correlates with number of integrations over the range of 1 to 8 integrations. Transgene expression decreased when a cell line with 15 integrations was analyzed. As a result, Mathor's statement that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell" is valid with respect to the range of 1 to 8 integrations and does not apply outside of that range. The Examiner's attempt to apply the statement outside of the range is not factually supported, i.e., supported by the data.

6. None of the other references relied on by the Examiner teach a correlation of transgene expression to integration number in the claimed range. Schott et al. teaches a correlation over the range of 1 to 9 integrations. See p. 304, Fig. 9. The increase between 4 and 5 integrations and 6 and 7 integration is much greater than the increase between 7 and 9 integrations. This is similar to the Mathor et al. data and indicates that transgene expression levels off as opposed to continuing to increase, although, as explained in more detail below, predications outside of the data range cannot be validly made.

7. I have addressed the data in Mathor et al. in my previous Declaration. The Examiner addresses the previous Declaration at pages 13-14 of the Office Action. The Examiner states:

> The applicant argues that the fact that different clones can produce different amounts of protein has no relevance to whether a person of skill in the art would modify Mathor and make clones with 20 or more integrated retroviral vectors. This is not found persuasive. On the contrary, such knowledge in the art does have relevance to whether one of skill in the art would make clones with 20 or more integrated retroviral vectors. The argument that the data in Table 1 of Mathor et al., which is limited to a maximum of 15 integrations cannot be extrapolated to a situation where there are 20 integrations is just an argument not supported by any evidence. Based on the teachings in the prior art (including Liu, Stamps, and Mathor et al.), one of skill in the art would have known that protein production is proportional to the number of integrated copies and that retroviral insertion is random and that expression level is dependent on the insertion sites; therefore, one of skill in the art would not conclude that the data in Table 1 indicates a maximum of 15 integrations. Based on the teachings in the art as a whole, one of skill in the art would have had reasonably expected that clones comprising more than 15 integrations would express higher amounts of protein and would have known to look for several clones having higher integration numbers and select the high producer clones.

The Examiner states that the argument that the data in Table 1 of Mathor et al., which is limited to a maximum of 15 integrations cannot be extrapolated to a situation where there are 20 integrations is just an argument not supported by any evidence. This is not true. First, the data in Table 1 of Mathor et al. show that a clone with 15 integrations has a lower level of expression than a clone with 8 integrations. This is a fact. It is also a fact that the experiments in Mathor et al. were not conducted in a manner so that a statistical analysis could be conducted. The groups were not replicated and there is no way to determine experimental error. Thus, it is not possible to construct a curve or equation from the data so that a correlation of transgene expression to a number of integrations outside of the data range (i.e., 20 to 100 integrations) can be made. Any attempt to do so is speculation without a factual basis. For example, based on the data in Mathor, it is speculation as to whether another clone with 15 integrations would have a level of transgene expression that is higher or lower than the reported clone. The reason for this is that the data is not amenable to statistical analysis so that such a prediction can be made.

This is the reason why the Examiner is incorrect in arguing that "one of skill in the art would have had reasonably expected that clones comprising more than 15 integrations would express higher amounts of protein and would have known to look for several clones having higher integration numbers and select the high producer clones." The data in Mathor et al. is not amenable to this assumption. Because the experiments were not replicated and because there is not data on multiple clones with 15 (or a similar number) of integrations, it is not possible to predict or comment on the amount of expression one could expect from another clone with 15 integrations. The single clone reported in Mathor et al. could be an example of the upper limit of expression, the median level of expression, or a low level of expression. The fact is, without additional data, one cannot know whether other clones with 15 integrations would have higher levels of expression than the observed clone. All that one knows for sure is the fact that the Mathor et al. data shows that expression from the clone with 15 integrations was lower than the expression observed in the clone with 8 integrations.

The Examiner further states that based on the teachings in the prior art (including Liu, Stamps, and Mathor et al.), one of skill in the art would have known that protein production is proportional to the number of integrated copies and that retroviral insertion is random and that expression level is dependent on the insertion sites; therefore, one of skill in the art would not conclude that the data in Table 1 indicates a maximum of 15 integrations.

As explained above, the data in Mathor et al. indicates that transgene expression increases with increases number of integrations up to 8 integrations. Transgene expression decreased when the number of integrations increased to 15. No other conclusions can be made based on this data. Furthermore, Lui et al. contains data on the correlation of expression of transgenes separated by an IRES and does not address transgene expression correlated to number of integrations. See Abstract, Fig. 2, Fig. 4, Fig. 5. Stamps et al. examined the role of the T-antigen gene and its site of integration in human epithelial cell immortalization. p. 871, Col. 2, first full para. The cells examined had up to five integrations. See Fig. 2 and Fig. 3. Stamps et al. does not comment on a correlation of transgene expression to number of integrations. These references do not provide factual support for the Examiner's argument.

8. In my previous Declaration, I provided evidence showing that at the time of the invention, the state of the art was that methylation of integrated retroviral vectors posed serious limitations on the use of the vectors for expression of transgenes. In response, the Examiner states:

> The applicant argues that many of the references cited by Bestor and those included in Paragraph 4 of the fourth Declaration describe silencing in vitro due to methylation. This is not found persuasive for the same reasons as above. Specifically, the prior art teaches that methylation is dependent on the integration site, i.e., consistent with the teachings of Liu, Stamps, and Mathor et al. that expression level is dependent on the insertion sites. Gunzburg et al. (The EMBO Journal, 1984, 3: 1129-1135) teach that retroviral integration is random and take place either in active (i.e., the virus is expressed) or in inactive (i.e., the virus is not expressed) chromatin domains (see p. 1129, paragraph bridging columns 1 and 2, p. 1133, column 2, p. 1134, column 1). Based on these teachings, one of skill in the art would have known that the same number of integrations would result in different expression levels, depending on the insertion site. Furthermore, the prior art teaches that the expression and stability of the gene of interest directly correlates with the number of integrated retroviral vectors (see Schott et al. above). One of skill in the art would have known to look for clones comprising high numbers of integrated retroviral vectors and select the ones capable of producing high amounts of protein.

Gunzburg et al., which was published in 1984, addresses methylation of "multiple endogenous mouse mammary tumour virus (MMTV) proviral genes" that "are present at different locations in mouse inbred strains." See Abstract. Gunzburg et al. finds that the methylation patterns are location and tissue specific and that the patterns are stably inherited and appear to be conferred upon the viral DNA by the flanking mouse genomic DNA. See Abstract. The authors state that "upon integration the provirus assumes the methylation pattern of the DNA into which is integrates." p. 1129, col. 1., p. 133, col. 2. Importantly, Gunzburg et al. does not contain data or comments that address any correlation of methylation to expression of genes. Just as important, Gunzburg et al. addresses endogenous proviral sequences and not the introduction of exogenous vectors containing transgenes. These proviral sequences are endogenous to the genome and have been acquired at some point in the distant past. Gunzburg et al. has very little relevance to the present invention or to the evidence I previously submitted.

Scientific papers that are more contemporary to this invention do contain relevant evidence.

Bestor and Tycko 1996 (attached at Tab 1), identify two hypothetical roles of genomic methylation patterns. The first is a role of programmed demethylation and methylation during development. p. 363, col. 1. The second role is that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA (i.e., integrated retroviruses). p. 363, col. 1.

This second role of methylation is directly relevant to the present invention which utilizes high levels of integrated retroviral vectors. Bestor and Tycko explain this relevance:

> The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features, and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome. Recognition and de novo methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The de novo sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted-repeats in supercoiled-plasmids are especially favored targets, as are secondary structures in artificial oligonucleotide substrates. This biochemical property suggests that invasive sequences might be targeted for de novo methylation because of their presentation of alternative secondary structures during integration (Fig la).

(p. 364, col. 2, Citations omitted).

Thus, it was a concern that due to the nature of retroviral integration, the retroviral vectors would be targeted for inactivation by methylation. Increasing copy number enhances this problem. "A common characteristic of invasive sequences is their presence in multiple copies, and it has

recently become known that repeated sequences can interact so as to trigger their mutual silencing." p. 364, col. 2. Bestor and Tycko further address retroviral vectors:

> Retroviral vectors that transducer reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals. Susceptability to de novo methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long term somatic gene therapy in humans. Successful gene transfer may require development of delivery vectors that evade the silencing response.

p. 365, col. 2.

Bestor and Tycko 1996 demonstrates why Gunzburg et al. is not relevant to the invention. Gunzburg et al. does not address the host defense mechanism at all or that fact that vectors had been shown to be actively silenced by methylation.

Garrick et al. 1998 (Tab 2) provide evidence on repeat-induced gene silencing in mammals. They used a lox/cre system to analyze the effect of copy number on transgene expression. They found that "reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus." p. 56, col. 1., p. 58. Again, this paper provides evidence that the state of the art was that increasing copy number leads to methylation and inactivation of transgenes. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene. As Bestor and Tycko 1996 indicated, the host defense mechanism is triggered by multiple copies of invasive sequences. The transposon system used in this paper and retroviral vectors are both invasive sequences.

Cherry et al. 2000 (Tab 3) is co-authored by two of the leading scientists in the field, Dr. David Baltimore and Dr. Rudy Jaenisch. They also recognize the role of methylation in the inactivation of proviral genes. They state: "DNA methylation is thought to be a general mechanism

used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genome wide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in Dnmtl knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types." p. 7419, col. 1-2. Both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression." p. 7425, col. 1.

Other papers also provide evidence regarding methylation and inactivation of multiple introduced copies of exogenous genes and inaction of retroviral vectors. Mehtali et al. 1990 (Tab 4) conducted experiments that show that methylation of an introduced transgene increases with increasing copy number and that expression of the transgene decreases with increasing copy number after initially increasing. See Table 1, p. 182. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene.

Niwa et al. 1983 (Tab 5) postulated that that there are two independent mechanisms that block expression from newly acquired retroviral vectors. See Abstract, p. 1105. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogenously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated. This paper further serves to demonstrate why the Examiner's reliance on Gunzburg et al. is inappropriate. Newly acquired retroviral vectors are treated by cells in a different manner from proviral sequences that have been integrated into the genome in the distant past and essentially become endogenous.

Svoboda et al. 2000 (Tab 6) examines the expression of retroviral vectors in foreign species. The vectors are subject to cell-mediated control at the transcriptional and posttranscriptional levels. Abstract, p. 181. Of main importance is cell transcriptional regulation, which can lead to proviral silencing. p. 181, col. 2. The authors note that all of the data so far point to the

important role of methylation in provirus silencing in general and that strategies for preventing methylation should contribute to more efficient gene transfer in the future. p. 186, col. 2. Again, the state of the art was that newly acquired retroviral vectors are subject to silencing by methylation. This is in direct contrast to the Examiner's conclusions based on Gunzburg et al.

Ellis and Pannell 2001 (Tab 7) also examine retrovirus silencing. They state that inclusion of appropriate regulatory elements may not be sufficient because the vectors are frequently silenced and that a better understanding of the mechanism of vector silencing is needed. p. 17, col. 1-2.

Challita and Kohn 1994 (Tab 8) provide data that shows that lack of expression following retroviral transduction is due to methylation. As stated by the authors: "Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by de novo methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts in vitro (9) and in vivo (2)." As shown by Ellis and Pannell (Tab 6), these problems still had not been solved by 2001, even when regulatory elements other than the retroviral LTR are used.

9. The references cited above establish that at the time of our invention the state of the art was that: 1) cells have a host defense mechanism that inactivates newly introduced, invading sequences such as retroviral vectors; 2) the host defense mechanism operates by methylation of the invading sequences, which causes transcriptional inactivation of the sequences; 3) transcriptional inactivation by methylation leads to reduced expression from retroviral vectors; 4) the inactivation may be triggered by structures formed during integration of the retroviral vectors; and 5) the presence of multiple repeats of an invading sequence such as a retroviral vector triggers methylation and inactivation.

10. The citations for the references cited above follow. These references are provided in the Appendix attached to this Declaration.

BESTOR TH et al, "Creation of genomic methylation patterns," Nature Genetics, 1996, Vol. 12(4) P. 363-7 (TAB 1)

GARRICK et al, "Repeat-induced gene silencing in mammals," Nature Genetics, Jan. 1998, Vol. 18(1), P. 56-9 (TAB 2)

CHERRY et al, "Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells," Molecular and Cellular Biology, Oct. 2000, Vol. 20(20) P. 7419-7426 (TAB 3)

METHTALI et al, "The methylation-free status of a housekeeping transgene is lost at high copy number," Gene, 1990, Vol. 91(2), P. 179-84 (TAB 4)

NIWA et al, "Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells," Cell, Apr. 1983, Vol. 32, P. 1105-1113 (TAB 5)

SVOBODA et al, "Retroviruses in foreign species and the problem of provirus silencing," Gene, 2000, Vol. 261, P. 181-188 (TAB 6)

ELLIS et al, "The beta-globin locus control region versus gene therapy vectors: a struggle for Expression" Clinical Genetics, Jan. 2001, Vol. 59(1) P. 17-24 (TAB 7)

CHALLITA et al, "Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo," Proc. Natl. Acad. Sci. USA, March 1994, Vol. 91, P. 2567-2571 (TAB 8)

11. I further declare that all statement made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Dated: September 14, 2010

Dr. Gregory Bleck

TAB 1

Creation of genomic methylation patterns

Timothy H. Bestor[1] & Benjamin Tycko[2]

There are two biological properties of genomic methylation patterns that can be regarded as established. First, methylation of 5'-CpG-3' dinucleotides within promoters represses transcription, often to undetectable levels. Second, in most cases methylation patterns are subject to clonal inheritance. These properties suit methylation patterns for a number of biological roles, although none of the current hypotheses can be regarded as proved or disproved. One hypothesis suggests that the activity of parasitic sequence elements is repressed by selective methylation. Features of invasive sequences that might allow their identification and inactivation are discussed in terms of the genome defense hypothesis. Identification of the cues that direct *de novo* methylation may reveal the biological role (or roles) of genomic methylation patterns.

[1]*Department of Genetics and Development*
[2]*Department of Pathology, Columbia University, College of Physicians and Surgeons, 701 W. 168th Street New York, New York 10032, USA*

Correspondence should be addressed to T.H.B.

The mammalian genome is ornamented with ~3×10^7 methyl groups, all at the 5 position of cytosine (m^5C) and most at 5'-CpG-3' dinucleotides. Methylation patterns increase the information content of the genome[1] and are transmitted by clonal inheritance[2]; methylation of CpG sites within promoters represses transcription[3]. This natural modification is also dangerous: m^5C is the major endogenous mutagen (deamination results in C→T transition mutations at CpG sites, which account for about one-third of all mutations in humans)[4], and tumour suppressor genes are frequently inactivated by ectopic *de novo* methylation of promoter regions[5,6]. However, there must be benefits that yield a net selective advantage. This is shown by the retention of cytosine methylation by virtually all organisms with genomes >5×10^8 basepairs[7], and by the demonstration that perturbations of methylation patterns are lethal to mouse embryos and to differentiated cells[8,9]. While methylation patterns clearly provide some essential function, the nature of that function or functions is still enigmatic.

There are at present two salient hypothetical roles of genomic methylation patterns. The first hypothesis (put forward twenty years ago) suggests that programmed demethylation and *de novo* methylation play a direct role in gene control during development[10,11]. Methylation patterns would be established during gametogenesis or early development, and regulatory factors would mediate the removal of methyl groups from promoters to allow the expression of tissue-specific genes at the appropriate stage of differentiation. While supported by a large body of indirect and correlative evidence, a definitive test of the causality of cytosine methylation in developmental gene control has been elusive. By the same token, a developmental role for cytosine methylation has not yet been disproved. It seems likely that methylation patterns might reinforce the heritability of states of gene expression mediated by chromatin proteins analogous to the Polycomb and trithorax group of proteins from *Drosophila*[12,13].

The second hypothesis suggests that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA[4,7]. Most of these elements are in fact methylated and transcriptionally inert in the genome of mammals, flowering plants, and those fungi whose genomes contain m^5C. Treatment of cultured cells or mice with the demethylating drug 5-azacytidine can activate silent retroviruses and endogenous genes that have been silenced by ectopic *de novo* methylation of regulatory regions[14]. It is striking that *Drosophila*, whose DNA lacks m^5C, suffers far larger numbers of insertion mutations than do animals whose genomes are methylated[15]. These observations, with support from evolutionary considerations[7], strongly suggest that cytosine methylation is part of a genomic host defense system that suppresses the transcription of parasitic sequence elements. The selective advantage of such a defensive system is obvious, given that a sizable fraction of the genome represents parasitic sequences that are invisible to the immune system and which might inflict intolerable mutational or cytotoxic damage if allowed to proliferate unchecked.

These hypothetical roles of genomic methylation in development or host defense place fundamentally different requirements on the DNA methylating system. The developmental role requires that methylation patterns be established as part of the developmental program via conventional sequence recognition mediated by sequence-specific DNA methyltransferases or specificity factors that interact with the ubiquitous DNA methyltransferase. The host-defense function requires a completely different

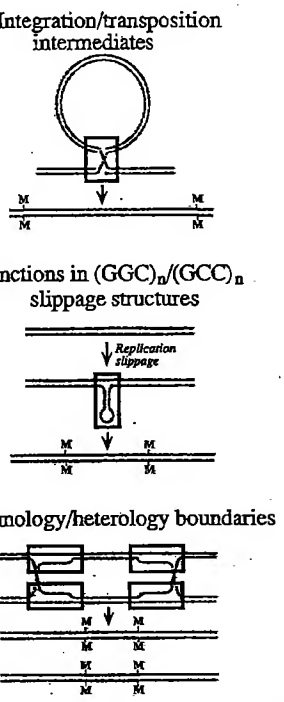

Fig. 1 Structure-dependent *de novo* methylation. Complementary sequences are depicted in identical colors, and structural features that are proposed to trigger silencing are boxed. M: 5-methylcytosine. *a*, A retrovirus integration intermediate showing the stage at which free 3' ends of the proviral DNA have been linked to 5' ends of chromosomal DNA. Most transposition and retroviral integration events utilize a common mechanism[19], and integrases and transposases share sequence or structural similarities. Recognition of integration/transposition intermediates by the silencing system could protect the genome from the proliferation of parasitic sequences. *b*, formation of three-way junctions in replication slippage structures at (GGC)/(GCC) repeats. Nearly all triplet repeat expansions at (GGC)/(GCC) sequences trigger *de novo* methylation in and around the repeats, and it is suggested that the three-way junction is the stimulus. Silencing of the *FMR1* gene in Fragile X syndrome is suggested to occur via this pathway. *c*, Detection of invasive sequences via homology/heterology boundaries during strand exchange[32]. Recognition and methylation of the boxed regions would result in reciprocal silencing of repeated sequences that reside in different sequence contexts.

source of specificity; *de novo* methylation must be cued by dynamic structural features unique to parasitic sequences in the act of invading the genome, or by events such as transient pairing interactions of repetitive elements. In other words, the developmental role requires that methylation be *directed* to specific sequences as part of the developmental programme, while the host defense role requires that methylation be *contingent* on the interaction of features characteristic of parasitic sequence elements with the DNA methylating system. Therefore it should be possible to deduce the role of methylation patterns from the nature of the biochemical events by which they were established.

Sequences that attract or repel *de novo* methylation in transfected embryonic cells have been identified[16–18], but many sequences (both foreign and endogenous) can be stably propagated in either the methylated or unmethylated state, and *de novo* methylation of such elements must therefore be dependent on features other than simple sequence recognition. The remainder of this article discusses possible mechanisms by which such sequences might be recognized and silenced.

Invasive DNA *in flagrante*

The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features[19], and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome[20]. Recognition and *de novo* methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The *de novo* sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted repeats in supercoiled plasmids are especially favored targets[21], as are secondary structures in artificial oligonucleotide substrates[22]. This biochemical property suggests that invasive sequences might be targeted for *de novo* methylation because of their presentation of alternative secondary structures during integration (Fig 1*a*).

Three-way junctions share structural features with four-way junctions, and enzymes that recognize four-way junctions (such as T4 endonuclease VII (ref. 23) and T7 endonuclease I (ref. 24)) recognize three-way junctions as well. The nature of the interaction of the methylating system with three-way junctions is of special interest because it may be involved in the aetiology of human diseases associated with GGC/GCC triplet repeat expansions. Slippage of replication intermediates can result in extrusion of a segment of GGC/GCC repeats, which form stable three-way junctions despite a lack of perfect complementarity in the extruded segment[22,25](Fig. 1*b*). The preference of DNA methyltransferase for junctions may result in *de novo* methylation of sequences in slippage structures. Diseases that result from GGC/GCC triplet repeat expansions may be thought of as autoimmune disorders, in which a host defense system attacks an innocuous alteration of an endogenous gene because of its incidental similarity to a parasitic sequence element in the act of integration.

Cytosine methylation and pairing interactions

A common characteristic of invasive sequences is their presence in multiple copies, and it has recently become known that repeated sequences can interact so as to trigger their mutual silencing. Fungi and flowering plants have diverse and highly effective means of silencing repeated sequences, and cytosine methylation is associated with silencing in nearly all cases. RIP (repeat-induced point mutation) imposes methylation, silencing, and large numbers of C→T transition mutations on repeated sequences during the sexual phase of the fungus *Neurospora crassa*[26], while MIP (methylation induced premeiotically) inactivates and methylates repeats during the sexual cycle of the fungus *Ascobolus immersus*[27]. Flowering plants can also silence and methylate repeated sequences via RIGS (repeat-induced gene silencing)[28,29]; the efficiency is such that multicopy transgenes frequently cause mutual silencing of themselves and of homologous resident sequences. The process can be remarkably efficient; unlinked transgene sequences as small as 300 bp can identify and inactivate each other in a genome of $>10^9$ bp.

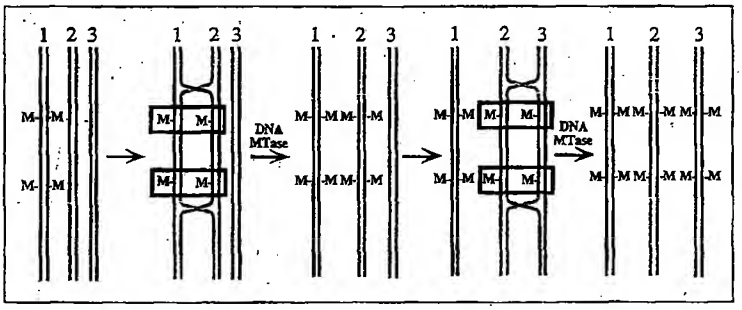

Fig. 2 Unidirectional transfer of epigenetic information via paired hemimethylated intermediates. Strand exchange between methylated and unmethylated repeated sequences presents hemimethylated sites, which are the preferred substrate of DNA methyltransferase; note that the maintenance activity of DNA methyltransferase leads to *de novo* methylation under these conditions. A common methylation pattern could propagate through an array or network of repeats by this mechanism, and methylation patterns could be transferred between alleles at loci that bear allele-specific methylation patterns.

The silenced state can persist even after the repeated sequences have been separated by segregation in sexual crosses[28]. Transgene silencing of this type is emerging as a barrier to the improvement of commercially-important plant species[30,31]. It is difficult to imagine a selective advantage for RIP, MIP, or RIGS other than host defense against the proliferation of parasitic sequence elements.

Transient pairing interactions are likely to be involved in the imposition of silencing and methylation on repeats; two mechanisms can be envisioned. First, strand exchange between repeats at non-allelic positions will present an abrupt loss of homology at the junction with flanking sequences (Fig. 1*c*), and silencing may be provoked by factors that recognize characteristic structural features at the homology/heterology boundary. This idea was developed by Signer and colleagues[32] to explain copy number-dependent transgene silencing in *Arabidopsis*. Second, methylation might propagate through a network of repeated sequences once one or more copies are methylated (Fig. 2). Strand exchange between methylated and unmethylated repeated sequences creates hemimethylated intermediates, which as mentioned previously, are strongly preferred substrates of DNA methyltransferase[12]; the preference for hemimethylated substrates normally contributes to maintenance methylation, but under these conditions *de novo* methylation is the result. Strand exchange reactions (which are thought to occur frequently as part of the double strand break repair pathway[33]) (Fig. 2) could cause a common methylation pattern to propagate through a network of repeats[28,30], and the probability of pairing with a methylated repeat will increase in proportion to the number of methylated repeats. A large number of methylated resident repeats will therefore increase the probability that a new, unmethylated copy will be methylated soon after its insertion into the genome. A large burden of parasitic sequences may have the paradoxical effect of repressing their own activity and that of homologous invasive sequences. Natural selection may actually favor the retention of large numbers of inactive parasitic sequences for this reason.

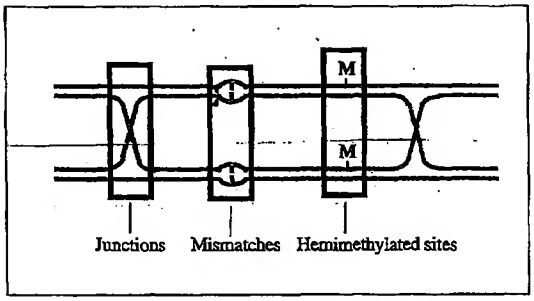

Fig. 3 Vulnerability of heteroduplex recombination intermediates to *de novo* methylation. Crossing-over during meiosis presents DNA methyltransferase with highly preferred targets: four-way junctions, mismatches and hemimethylated sites. Dysregulated *de novo* methylation appears to be prevented by sharp down-regulation of DNA methyltransferase during the pachytene stage of meiosis; the mechanism involves the production of a larger, non-translated DNA methyltransferase mRNA[48].

Many endogenous genes are present in multiple copies which escape repeat-dependent gene silencing. In *Neurospora*, which seems to be especially aggressive in its response towards repeated sequences, the efficiency of RIP is greater when the repeats are in close proximity[26]. This argues for special protective mechanisms that shield repeated cellular genes from repeat-induced silencing. A complete intolerance of repeated sequences would also put severe constraints on the evolution of new functions, which depends on the duplication and divergence of existing genes. However, the factors that control sensitivity to repeat-dependent silencing are poorly understood[34].

There is no clear experimental evidence that repeat-dependent silencing or methylation operates in mammals, as is the case in fungi and plants. However, no objective search for repeat-dependent silencing in mammals seems to have been conducted. Circumstantial evidence suggests that it may occur[35]. Much of the m^5C in the mammalian genome is found in repetitive DNA, and most of the repeated and potentially transposable elements in the genome are methylated and transcriptionally inert through most of development[36,37]. Retroviral vectors that transduce reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals[38]. Susceptibility to *de novo* methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long-term somatic gene therapy in humans. Successful gene transfer may require the development of delivery vectors that evade the silencing response. It is also possible that the development of therapeutic agents that perform selective methylation of deleterious sequence elements (such as HIV-1 proviral DNA) will activate an existing host defense system and therefore serve as a sort of nuclear vaccine.

Maintenance of allele-specific methylation patterns

Mammalian genes whose expression is dependent on the sex of the contributing parent are said to be imprinted; such genes bear allele-specific methylation patterns that are necessary for maintenance of the imprinted state[9]. Short repetitive sequences are characteristic of imprinted genes[39,40]; these short repeats may form alternative secondary structures that trigger *de novo* methylation in the germ line or during early postfertilization development. A parent-of-origin effect on gene expression results if there

are differences in the types of DNA or chromatin structures formed, or if the response to the structures differs, in the male and female germlines. In this way a nuclear host defense system may give rise to allele-specific methylation patterns and to the genetic phenomenon of imprinting[41], and could also account for documented gamete-of-origin effects on methylation of repetitive DNA[37]. It should be noted that several mutations that result from insertion of retroviruses of the intracisternal A particle (IAP) type confer imprinted behavior on genes that are not imprinted in the wild type[42,43].

Alleles at imprinted loci are asymmetrically methylated, and the mechanism shown in Fig. 2 will tend to convert asymmetrical allelic methylation patterns towards the more heavily methylated pattern. Certain pathological human conditions show abnormalities in the functional imprinting of particular chromosomal regions which might arise via this type of interchromosomal transfer of epigenetic information. For example, Wilms' tumours frequently show conversion to a uniparental (bipaternal) methylation and expression pattern at imprinted loci in the *H19/IGF2* region on chromosome 11p15.5, which could result from the local transfer of methylation patterns from the paternal chromosome to the less heavily-methylated maternal chromosome[44]. It is important to note, however, that a conversion to a symmetrical uniparental allelic methylation pattern at imprinted loci could also come about by an alternative mechanism that involves deletion or mutations of a *cis*-acting DNA element (an 'imprinting centre'). This has been implicated in certain kindreds that show disruption of imprinting patterns within the chromosome 15 Prader Willi/Angelman syndrome region[45].

Vulnerability of meiotic recombination intermediates

Allelic heteroduplex recombination intermediates present several features that should make them very vulnerable to *de novo* methylation (Fig. 3). First, allelic methylation differences create hemimethylated sites in the heteroduplexes, which provoke *de novo* methylation via the maintenance activity of DNA methyltransferase (see Fig. 2). Second, Holliday structures and four-way junctions are necessarily present; as described earlier, these structures appear to be favored targets of DNA methyltransferase. Third, mismatches in the vicinity of CpG dinucleotides greatly favor *de novo* methylation, presumably by lowering the energetic barrier associated with eversion of the target cytosine during the transmethylation reaction[46,47]. Presentation of these vulnerable sites might create a "methylation ratchet" in which methylation levels increase in an unregulated fashion with each meiotic cycle.

Examination of purified germ cells from male mice showed that both DNA methyltransferase protein and the 5.2-kb DNA methyltransferase mRNA found in all proliferating cell types was present in all germ cell fractions, except for pachytene spermatocytes. These spermatocytes showed an absence of DNA methyltransferase protein and contained a 6.2-kb RNA that was not associated with polyribosomes. It is the pachytene stage of meiosis I where most crossing-over occurs. The ubiquitous 5.2-kb DNA methyltransferase mRNA and DNA methyltransferase protein reappeared at the conclusion of the crossing-over phase of meiosis[48]. These findings suggest that meiotic recombination intermediates are protected from *de novo* methylation through down-regulation of DNA methyltransferase via a novel post-transcriptional mechanism that involves the production of a larger, non-translated RNA transcript.

The function of cytosine methylation

The fact that cytosine methylation can increase the information content of DNA has tempted many to attribute diverse roles to methylation patterns. Cytosine methylation has been proposed to reduce the effective size of the genome by masking non-regulatory regions in large-genome eukaryotes[7,49], and central roles in DNA repair[22,50] and replication[51] have also been mooted. None of the hypothetical functions of cytosine methylation (and this includes the developmental and host-defense functions) has the support of compelling experimental evidence, and all, some, or none of the hypotheses may be valid. At this time it seems that the true function of cytosine methylation will be understood only when we learn how the cell selects specific sequences for covalent modification.

Aknowledgements

E. Selker and M. Turker made helpful comments on the manuscript. Supported by NIH grants GM00616 and CA60610 (to T. H. B.) and NIH grant CA60765 and American Cancer Society grant JFRA482 (to B.T.).

Received 27 November 1995; accepted 3 January 1996.

1. Reinisch, K. M., Chen, L., Verdine , G.L. & Lipscomb, W.N. (1995) The crystal structure of *HaeIII* methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* 82, 143–153 (1995).
2. Pfeifer, G., Steigerwald, S., Hansen, R.S. Gartler, S.M. & Riggs, A.D. (1990) Polymerase-chain reaction aided genomic sequencing of an X chromosome linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA* 87, 8252–8256 (1990).
3. Iguchi-Ariga, S.M.M. & Schaffner, W. CpG methylation of the cAMP responsive enhancer/promoter TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* 3, 612–619 (1989).
4. Bestor, T.H. & Coxon, A. The pros and cons of cytosine methylation *Curr. Biol.* 3, 384–386 (1993).
5. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl. Acad. Sci. USA* 92, 7416–7419 (1995).
6. Herman, J. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA* 91, 9700–9704 (1994).
7. Bestor, T.H. DNA methylation: evolution of a bacterial immune function into a regulator of genome structure and activity in higher eukaryotes. *Phil. Trans. Royal Soc. Lond.* B 326, 179–187 (1990).
8. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of mouse DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926 (1992).
9. Li, E., Beard, C., Forster, A.C., Bestor, T.H. & Jaenisch, R. DNA methylation, genomic imprinting, and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* LVIII, 297–305 (1993).
10. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* 187, 226–232 (1975).
11. Riggs, A.D. X-inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 14, 9–11 (1975).
12. Bestor, T.H. & Verdine, G.L. DNA methyltransferases. *Curr. Opin. Cell Biol.* 6, 380–389 (1994)
13. Bestor, T.H. The biochemical basis of allele-specific gene expression in genomic imprinting and X inactivation. in *Genomic imprinting: Causes and consequences.* (eds Ohlsson, R., Hall, K. & Ritzen, M.) (Cambridge University Press, Cambridge, 1995).
14. Jaenisch, R., Schnieke, A. & Harbers, K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc. Natl. Acad. Sci. USA* 82, 1451–1455 (1985).
15. Ashburner, M. *Drosophila: A Laboratory Handbook.* 303 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992).
16. Mummaneni, P., Bishop, P. & Turker, M. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. *J. Biol. Chem.* 268, 552–558 (1993).
17. Macleod, D., Charlton, J., Mullins, J. & Bird, A. Sp1 binding sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* 8, 2282–2292 (1994)
18. Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* 371, 435–438 (1994)
19. Craig, N.L. Unity in transposition reactions. *Science* 270, 253–255 (1995).
20. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371, 215–220 (1994).
21. Bestor, T.H. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. *Nucl. Acids Res.* 15, 3835–3843 (1987).
22. Chen, X. *et al.* Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* 92, 5199–5203 (1995).
23. Jentsch, F. & Kemper, B. Endonuclease VII resolves Y-junctions in branched DNA *in vitro*. *EMBO J.* 5, 181–190 (1986).
24. Lu, M., Guo, Q., Studier, W.F. & Kallenbach, N.F. Resolution of branched DNA substrates by T7 endonuclease I and its inhibition. *J. Biol. Chem.* 266, 2531–2536 (1991).
25. Usdin, K. & Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucl. Acids Res.* 23, 4202–4209 (1995).
26. Selker, E.U. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* 24, 579–613. (1990).
27. Goyon, C., Nogueira, T.I.V. & Faugeron, G. Perpetuation of cytosine methylation in *Ascobolus immersus* implies a novel type of maintenance methylase. *J. Mol. Biol.* 240, 42–51 (1994).
28. Matzke, M.A. & Matzke, A.J.M. Homology-dependent gene silencing in transgenic plants: what does it really tell us? *Trends Genet.* 11, 1–3 (1995).
29. Matzke, A.J.M., Neuhuber, F., Park, Y.D., Ambros, P.F. & Matzke, M.A. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes *Mol. Gen. Genet.* 244, 219–229 (1994).
30. Flavell, R.B. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl. Acad. Sci. USA* 91, 3490–3496 (1994).
31. Finnegan, J. & McElroy, D. Transgene inactivation: plants fight back! *Biotechnol.* 12, 883–888. (1994)
32. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* 22, 1067–1085 (1993).
33. Engels, W.R., Preston, C.R. & Johnson-Schlitz, D.M. Long-range cis preference in DNA homology search over the length of the Drosophila chromosome. *Science* 263, 1623–1625 (1994).
34. Bender, J. & Fink, G. Epigenetic control of endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* 83, 725–734 (1995)
35. Kricker, M.C., Drake, J.W. & Radman, M. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc. Natl. Acad. Sci. USA* 89, 1075–1079 (1992).
36. Chesnokov, I.N. & Schmid, C.W. Specific Alu binding protein from human sperm chromatin prevents DNA methylation. *J. Biol. Chem.* 270, 18539–18542 (1995)
37. Rubin, C.M., VandeVoort, C.A., Tiplitz, R.L. & Schmid, C.W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucl. Acids Res.* 22, 5121–5127 (1994).
38. Challita, P.-A. & Kohn, D.B. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc. Natl. Acad. Sci. USA* 91, 2567–2571 (1994).
39. Neumann, B., Kubicka, P. & Barlow, D.P. Characteristics of imprinted genes. *Nature Genet.* 9, 451–452 (1994).
40. Hatada, I. & Mukai, T. Genomic imprinting of *p57KIP2*, a cyclin-dependent kinase inhibitor in mouse. *Nature Genet.* 11, 204–206 (1995).
41. Barlow, D.P. Methylation and imprinting: from host defense to gene regulation? *Science* 260, 309–310 (1993).
42. Michaud, E. *et al.* Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* 8, 1463–1472 (1994).
43. Duhl, D., Vrieling, H., Miller, K., Wolff, G. & Barsh, G. Neomorphic *agouti* mutations in obese yellow mice. *Nature Genet.* 7, 220–226 (1994).
44. Moulton, T. *et al.* Epigenetic lesions at the *H19* locus in Wilms' tumour patients. *Nature Genet.* 7, 440–447 (1994).
45. Buiting, K. *et al.* Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature Genet.* 9, 395–400 (1995).
46. Yang, A.S., Shen, J.-C., Zingg, J.-M., Mi, S. & Jones, P.A. HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucl. Acids Res.* 23, 1380–1387 (1995).
47. Klimasauskas, S. & Roberts, R.J., M.HhaI binds tightly to substrates containing mismatches at the target base. *Nucl. Acids Res.* 23, 1388–1395 (1995).
48. Jue, K., Bestor, T.H. & Trasler, J. Regulated synthesis and localization of DNA methyltransferase during spermatogenesis. *Biol. Reprod.* 53, 561–569 (1995).
49. Bird, A.P. Gene number, noise reduction and biological complexity. *Trends Genet.* 11, 94–100 (1995).
50. Cleaver, J. It was a very good year for DNA repair. *Cell* 76, 1–4 (1994)
51. Taylor, J.H. Origins of replication and gene regulation. *Mol. Cell. Biochem.* 61, 99–109 (1984).

TAB 2

Repeat-induced gene silencing in mammals

David Garrick[1], Steven Fiering[2,3], David I.K. Martin[2] & Emma Whitelaw[1]

In both plants[1–3] and *Drosophila melanogaster*[4,5], expression from a transgenic locus may be silenced when repeated transgene copies are arranged as a concatameric array. This repeat-induced gene silencing is frequently manifested as a decrease in the proportion of cells that express the transgene, resulting in a variegated pattern of expression. There is also some indication that, in transgenic mammals, the number of transgene copies within an array can exert a repressive influence on expression, with several mouse studies reporting a decrease in the level of expression per copy as copy number increases[6–8]. However, because these studies compare different sites of transgene integration as well as arrays with different numbers of copies, the expression levels observed may be subject to varying position effects as well as the influence of the multicopy array. Here we describe use of the lox/Cre system of site-specific recombination to generate transgenic mouse lines in which different numbers of a transgene are present at the same chromosomal location, thereby eliminating the contribution of position effects and allowing analysis of the effect of copy number alone on transgene silencing. Reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus. These findings establish that the presence of multiple homologous copies of a transgene within a concatameric array can have a repressive effect upon gene expression in mammalian systems.

The αPElox construct used to generate transgenic mouse lines is shown in Fig. 1*a*. In this transgene, expression of the *lacZ* reporter gene is driven by the human α-globin promoter and the αHS-40 enhancer-like element of the α-like globin locus[9]. As expression of the *lacZ* reporter can be analysed in single erythroid cells, use of this transgene allows the detection of variegated patterns of expression[10]. The transgene also contains a single copy of the 34-bp recognition site (*loxP*) of the Cre recombinase of bacteriophage P1 (ref. 11). Cre-mediated recombination between *loxP* sites in individual αPElox transgenes that form part of a concatameric array will reduce the transgene copy number without altering the site of integration[12] (Fig. 1*b*). Although transgenes within a multicopy array in mice are usually present in a tandem (head–tail) orientation, inverted repeats (head–head or tail–tail) do occur[13]. Because Cre recombination between inverted *loxP* sites causes the inversion rather than excision of the intervening DNA, only recombination between sites in like-oriented transgene monomers will reduce the copy number of the array.

We generated founder transgenic mice bearing the αPElox construct, and from them we established hemizygous transgenic lines. Southern blotting of tail DNA revealed that two of these transgenic lines (αPElox1 and αPElox7) contain more than 100 copies of the transgene (Fig. 2), and these lines were selected for Cre-mediated reduction in transgene copy number. For each of these parent lines, we collected fertilized oocytes from wild-type female mice mated with G_1 hemizygous transgenic males. Oocytes were micro-injected with a circular Cre expression vector (pCAGGS-Cre) and then transferred into foster mothers. Transient expression of the recombinase from the unincorporated pCAGGS-Cre plasmid has been shown to catalyze efficient site-specific recombination at *loxP* sites within the mouse genome before the morular stage of development[14]. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus. Cre recombination within the αPElox1 parent line (more than 100 copies) gave rise to two distinct reduced-copy progeny lines: 1.cre/a, which contains five copies of the transgene, and 1.cre/b, in which the array has been reduced to a single copy (Fig. 2). Southern-blot analysis indicated that the single copy remaining in the 1.cre/b line contains a rearrangement/deletion (data not shown). The αPElox7 parent line (more than 100 copies) gave rise to a single derivative line (7.cre/a) bearing one copy of the transgene.

For parental and reduced-copy progeny lines, transgene expression was analysed in 12.5-dpc embryos by staining of whole primitive erythrocytes with X-gal. We previously showed that all cells containing β-galactosidase activity can be detected by light microscopy after staining under these conditions[10,15]. For both of the high-copy parent lines, a heavily variegated pattern of transgene expression was observed, with less than 1% of

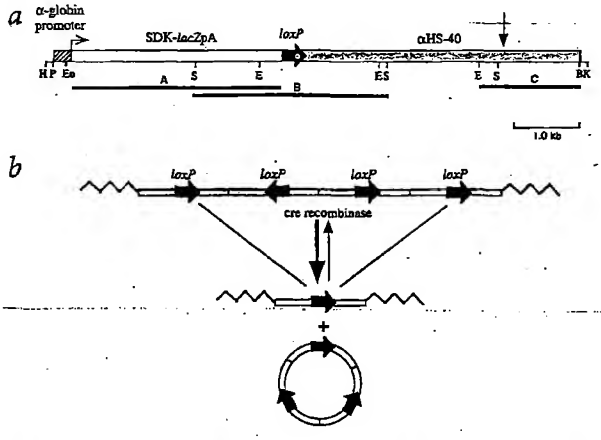

Fig. 1 Strategy for Cre-mediated reduction of transgene copy number. **a**, The αPElox transgene construct contains nucleotides −573 to +36 of the human α-globin promoter upstream of the SDK–*lacZ*pA cassette used previously[30] and a 4.2-kb fragment containing the αHS-40 DNaseI hypersensitive site (vertical arrow). An oligonucleotide containing a single copy of the 34-bp *loxP* site was inserted between the SDK–*lacZ*pA cassette and the αHS-40 fragment. Black lines indicate probes used in this study. H, *Hin*dIII; P, *Pst*I; Eo, *Eco*0109 I; S, *Sac*I; E, *Eco*RI; B, *Bam*HI; K, *Kpn*I. **b**, When multiple copies of the αPElox transgene are situated within a concatameric array in the mouse genome, Cre-mediated recombination between like-oriented *loxP* sites within the array removes the intervening DNA as a circular episome, leaving a reduced-copy array at the same chromosomal location. The forward (excision) reaction is heavily favoured over the reverse (integration) reaction, which requires recombination between *loxP* sites on separate molecules.

[1]*Department of Biochemistry, University of Sydney, New South Wales 2006, Australia.* [2]*Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.* [3]*Current address: Department of Microbiology, Dartmouth Medical Center, Lebanon, New Hampshire, USA. Correspondence should be addressed to E.W. e-mail: e.whitelaw@biochem.usyd.edu.au*

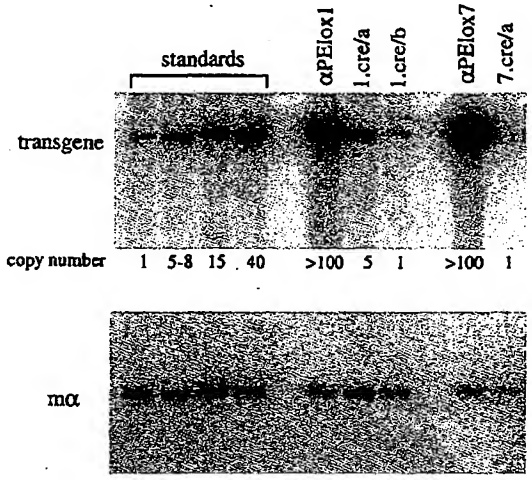

Fig. 2 Copy-number determination for αPElox parental and Cre-modified progeny transgenic lines. DNA was obtained from tail biopsies of three-week-old G_1 hemizygous mice for each of the lines shown, as well as from lines bearing known copy numbers of a transgene that contains the same *lacZ*–αHS-40 cassette. After digestion with *Sac*I, DNA was resolved on a 0.8% agarose gel and transferred to a nylon membrane. The membrane was hybridized with fragment B (Fig. 1*a*), and the copy number was determined by quantifying the intensity of the resulting signal and comparing it with those of the standards. To ensure equivalent DNA loadings in each lane, the membrane was stripped and re-hybridized with a probe for the murine α-globin gene (mα). In a separate Southern blot performed on tail DNA from each of these lines, hybridization with a probe specific for the 3′ end of the transgene (fragment C, Fig. 1*a*) revealed that the junction fragments between insert and mouse DNA generated with the restriction enzymes *Eco*RI or *Eco*RV for each of the reduced-copy progeny lines were the same as those observed in the respective parent arrays, indicating that the genomic location of the array had not been altered during Cre modification (data not shown).

erythroid cells containing an active transgene locus (Fig. 3, Table 1). Similar findings of variegated expression patterns in high-copy transgenic lines bearing a construct (αPE) that differs from αPElox only in the absence of the single *loxP* element[16] suggest that the presence of the 34-bp *loxP* oligonucleotide has no effect on transgene expression. When the number of transgene copies within the αPElox1 array was reduced to five by Cre-mediated recombination (1.cre/a), there was a large increase (more than 1,000-fold) in the percentage of primitive erythroid cells expressing the transgene. Similarly, Cre-mediated reduction of the αPElox7 parent array to one copy (7.cre/a) also suppressed the variegated expression that was observed in the parent line, with a 180-fold increase in the size of the expressing population (Fig. 3, Table 1). In these two independent transgenic lines, the presence of high-copy multimeric arrays is therefore associated with a silencing of transgene expression, which is observed as a decrease in the percentage of cells containing an active transgene locus. A decrease in the number of copies within each array correlates with a suppression of variegation—that is, an increase in the size of the expressing population. No expression of the transgene was observed in the line 1.cre/b, in which the single transgene copy was rearranged (data not shown).

In plants, repeat-induced gene silencing has been observed at both the transcriptional[17–19] and post-transcriptional[20,21] stages of gene expression. To determine whether the copy-number–dependent silencing of transgene expression in mouse erythroid cells occurs at transcription or involves a post-transcriptional modification, we performed run-on analysis in nuclei of 12.5-dpc erythrocytes from αPElox1 and its reduced-copy derivative, 1.cre/a (Fig. 4*a*). Run-on transcripts from the *lacZ* reporter gene were present in 1.cre/a primitive erythrocytes but were not detectable in erythroid cells of the high-copy parent line, indicating that the silencing of expression from multi-copy transgene arrays occurs at the level of transcription. The methylation status and local chromatin structure of the transgene locus in these two lines were also compared. Although transgenes present in 12.5-dpc erythrocytes of the reduced-copy line 1.cre/a appear unmethylated, the high-copy αPElox1 transgene locus at the same genomic location was found to be heavily methylated (Fig. 4*b*). To analyse chromatin structure, we performed endonuclease protec-

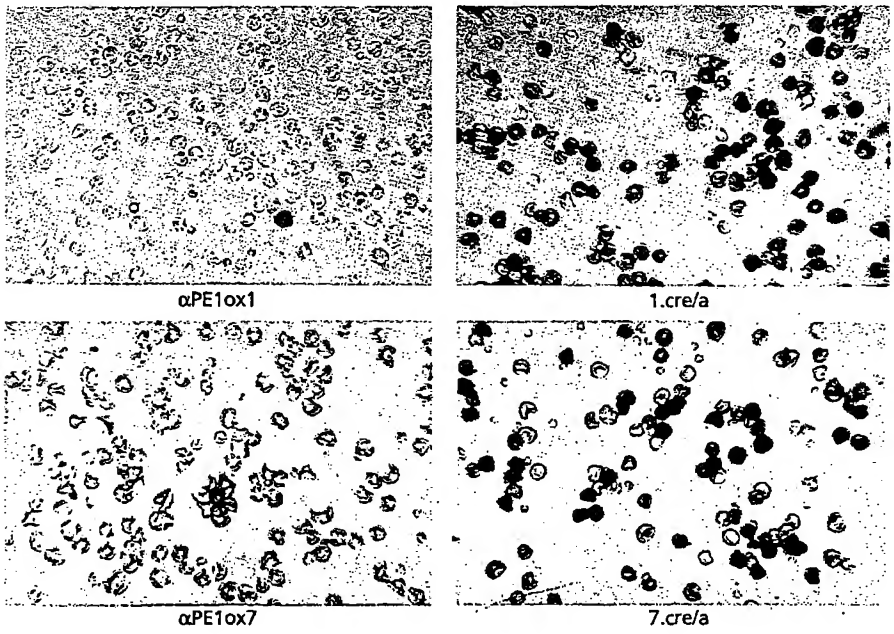

Fig. 3 X-gal staining of primitive erythrocytes from the αPElox transgenic lines and their reduced-copy progeny. Individual 12.5-dpc embryos from the transgenic lines shown were bled into PBS and whole blood cells were stained with X-gal before visualization under light microscopy. For αPElox1, many fields of view had to be scanned to detect a single blue cell.

Table 1 • Copy number and erythroid expression

Transgenic line	Copy number	Percentage of 12.5-dpc erythroid cells expressing *lacZ*
αPElox1	>100	0–0.01 (n=11)
1.cre/a	5	65±9 (n=11)
αPElox7	>100	0.3±0.02 (n=4)
7.cre/a	1	54±9 (n=3)

Transgene copy numbers were determined by Southern analysis (Fig. 2). The percentage of 12.5-dpc erythroid cells expressing *lacZ* was determined by staining cells as described (Fig.3) and scoring a minimum of 200 cells for a detectable blue colour. Expression data are presented as the mean ± one standard deviation—except for αPElox1, where a range is given. n, number of individual transgenic embryos assayed for each line.

tion assays[22] on nuclei isolated from 12.5-dpc primitive erythroid cells from these two lines. In their native chromatin configuration, transgenes in the high-copy array (αPElox1) were more resistant to endonuclease digestion at a site within the transgene promoter than transgenes that were at the same genomic location but within a lower-copy array (1.cre/a; Fig. 4*c*). The transcriptional silencing that occurs at high-copy arrays is therefore associated both with hypermethylation of transgenes in the locus and with the adoption of a repressive local chromatin configuration.

These results establish that the presence of multiple repeats within a high-copy array can directly repress transgene expression in a mammalian system. Because transgenic animals frequently contain high-copy concatameric arrays of a monomeric unit, this observation has important implications for the meaningful interpretation of transgenic experiments in mammals. The influence of copy number on silencing is likely to be most prominent in lines such as those studied here, in which the array is very large, but repeat-induced silencing may often be responsible for poor transgene expression. In addition to copy number, as shown here, the factors influencing transgene silencing in mice include integration site, the lineage in which expression occurs and the *cis*-acting control elements within the transgene[23]. Transgene constructs for which copy-number–dependent expression has been reported[24,25] may contain genetic elements that function to insulate individual monomers and prevent silencing. Without such elements, the lox/Cre system of site-specific recombination could be used to activate expression in high-copy lines.

The hypermethylation of transgenes within the inactive, high-copy array noted here has also been reported in other cases of repeat-induced silencing[1,3,26,27], although it is unclear in our system whether methylation precedes or is a consequence of the observed chromatin restructuring. The occurrence of copy-number–dependent transgene inactivation in *Drosophila*, in which DNA methylation has not been detected[28], suggests that methylation is not necessary for the chromatin remodelling associated with repeat-induced silencing and that hypermethylation of the αPElox transgene in mice may be a secondary modification that occurs at already inactive, high-copy arrays. This hypothesis is consistent with our previous finding that the variegated silencing of globin/*lacZ* transgenes in mice correlates with an inactive chromatin structure but not hypermethylation at the transgene locus[29].

Although a molecular mechanism for repeat-induced silencing of multimeric arrays is yet to be fully elucidated, the correlation between the silenced state and the adoption of a less accessible

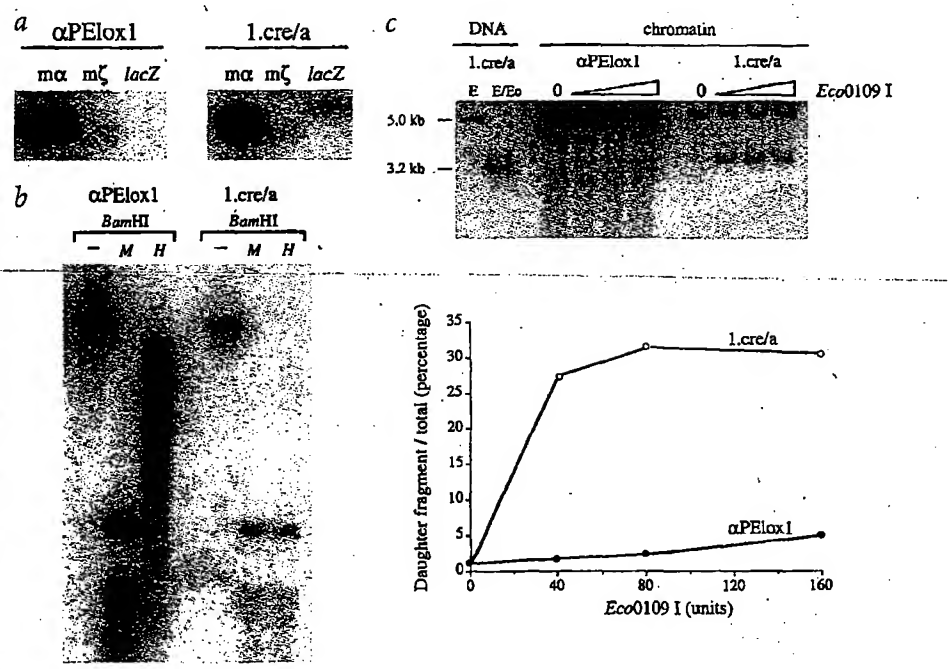

Fig. 4 Transcription, methylation and chromatin accessibility of the transgene. *a*, Nuclear run-on analysis of *lacZ* transcription in αPElox1 and 1.cre/a. Nuclei were prepared from primitive erythroid cells collected from 12.5-dpc transgenic embryos of the lines αPElox1 and 1.cre/a. Run-on transcripts were synthesized and hybridized to membranes containing the following DNA fragments: the mouse α-globin gene from +17 to +956 (mα), the mouse ζ-globin gene from +150 to +969 (mζ) and a fragment of the *lacZ* gene, which includes nucleotides from +39 to +1164 relative to the start site of the αPElox transcript (*lacZ*). *b*, Analysis of methylation status of the transgene locus in αPElox1 and 1.cre/a. DNA from 12.5-dpc erythrocytes of the lines αPElox1 and 1.cre/a was digested with either *Bam*HI alone or *Bam*HI together with either *Msp*I (M) or *Hpa*II (H), which are insensitive and sensitive, respectively, to methylation at the central CpG dinucleotide of their common recognition sequence, 5′–CCGG. The resulting fragments were analysed by Southern blotting and hybridization with fragment A (Fig. 1*a*). *c*, Endonuclease protection assay of the transgene locus in αPElox1 and 1.cre/a. Nuclei prepared from primitive erythroid cells of the lines αPElox1 or 1.cre/a were digested with increasing concentrations of *Eco*0109 I. The DNA was then purified and digested with *Eco*RI to release a 5.0-kb parent fragment (composed of the tail–head junction between adjacent transgenes) before Southern analysis (top panel). When the membrane was hybridized with fragment A (Fig. 1*a*), cleavage at the *Eco*0109 I site within the transgene promoter reduced this parent band to a 3.2-kb daughter fragment. The sizes of the expected parent and daughter fragments were determined by digestion of purified tail DNA from the line 1.cre/a with either *Eco*RI alone (E) or both *Eco*RI and *Eco*0109 I (E/Eo). The membrane was quantified with a PhosphorImager and Imagequant version 4.2a software (Molecular Dynamics; bottom panel). The vertical axis shows the percentage of total signal contained within the daughter fragment at each enzyme concentration for the lines αPElox1 (closed circles) and 1.cre/a (open circles).

chromatin configuration observed both here and in *Arabidopsis*[19] is consistent with a model in which homologous pairing between monomers within the array induces heterochromatinization at the transgene locus. Support for a model of heterochromatin formation is derived from the suppression of copy-number–dependent silencing of pigment genes in the *Drosophila* eye by mutations within genes encoding known structural components of heterochromatin[4]. It remains to be determined whether the repeat-induced modification of chromatin structure in mice is dependent on the proximity of the locus to nearby blocks of constitutive heterochromatin, as was the case for a *brown* eye pigment transgene in *Drosophila*[5], or whether high-copy arrays autonomously form inactive chromatin structures irrespective of their position. The arrangement of endogenous loci such as the rRNA, tRNA and histone genes as high-copy concatamers of a repeated unit suggests that multi-copy arrays need not always be subject to repeat-induced silencing. An investigation of how silencing is prevented at these endogenous loci may prove useful in maintaining activity at multi-copy arrays of foreign genetic elements.

Methods

Transgenic mice. The αPElox transgene was constructed by insertion of an oligonucleotide containing a single copy of the 34 bp *loxP* site (5′-ATAACTTCGTATAATGTATGCTA TACGAAGTTAT-3′) between the SDK–*lacZ*pA cassette and the 4.2-kb αHS-40 fragment of the previously described construct αPE[16]. The integrity of the *loxP* site was confirmed by dideoxy sequencing. The αPElox fragment for micro-injection was excised from plasmid vector sequences by digestion with *Hin*dIII and *Kpn*I and purified by agarose-gel electrophoresis. Transgenic mouse lines bearing the αPElox construct were generated by standard micro-injection techniques in the outbred Pathology Oxford (P.O.) mouse strain. For Cre recombination, fertilized oocytes were collected from wild-type P.O. females mated with hemizygous G_1 transgenic male mice and micro-injected into either pronucleus with the circular Cre-expression vector pCAGGS-Cre[14] at 5 ng/μl before transfer into pseudo-pregnant foster mothers. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus.

Histology. After individual 12.5-dpc embryos had been bled into PBS, whole blood cells were gently pelletted and then fixed in 0.25% glutaraldehyde before staining with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) at 37 °C for 24 h as described previously[30].

Nuclear run-on and endonuclease protection assays. To prepare membranes for nuclear run-on analysis, probe fragments were excised from plasmid vector by digestion with the appropriate restriction enzymes and purified by agarose-gel electrophoresis. Probe fragments used were an *Hin*fI–*Hin*fI fragment containing +17 to +956 relative to the transcription start site of the mouse α-globin gene; an *Xba*I–*Pst*I fragment containing +150 to +969 of the mouse ζ-globin gene; and a *Pst*I–*Eco*RV fragment of pSDK-*lacZ*pA[30], which includes +39 to +1164 relative to the start site of the *lacZ* transcript produced by αPElox. Purified fragment DNA (0.2 μg of each fragment) was then electrophoresed on 1.2% agarose gel and transferred to a nitrocellulose membrane by Southern blotting. Nuclei were purified as described[9] from 1.5×10^7 primitive erythroid cells collected from 12.5-dpc transgenic embryos. Run-on transcripts were synthesized from isolated nuclei and hybridized to membranes as described previously[31]. For endonuclease protection assays, nuclei prepared as described above were separated into four aliquots of 250 μl and digested with 0, 40, 80 or 160 U of *Eco*0109 I at 37 °C for 90 min. After proteinase-K digestion and phenol-chloroform extraction, purified DNA was digested with *Eco*RI and Southern analysis performed.

Acknowledgements

We thank P. Vassalli for the gift of the pCAGGS-Cre plasmid, H. Sutherland for the gift of the αPE plasmid, W. Magis for technical support and M. Groudine for support and helpful discussions. This work was supported by a grant from the National Health and Medical Research Council of Australia (to E.W.), an Australian Postgraduate Award (to D.G.), an NIH grant (to D.I.K.M.) and a grant from the Burroughs-Wellcome Fund to S.F. D.I.K.M. is a scholar of the Leukemia Society of America.

Received 6 August; accepted 10 October, 1997.

1. Linn, F., Heidmann, I., Saedler, H. & Meyer, P. Epigenetic changes in the expression of the maize A1 gene in *Petunia hybrida*: role of numbers of integrated gene copies and state of methylation. *Mol. Gen. Genet.* **222**, 329–336 (1990).
2. Scheid, O.M., Paszkowski, J. & Potrykus, I. Reversible inactivation of a transgene in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **228**, 104–112 (1991).
3. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).
4. Dorer, D.R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**, 1–20 (1994).
5. Sabl, J.F. & Henikoff, S. Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. *Genetics* **142**, 447–458 (1996).
6. Davis, B.P. & MacDonald, R.J. Limited transcription of rat elastase I transgene repeats in transgenic mice. *Genes Dev.* **2**, 13–22 (1988).
7. Dorer, D.R. Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res.* **6**, 3–10 (1997).
8. Sharpe, J.A. *et al.* Analysis of the human α-globin gene cluster in transgenic mice. *Proc. Natl. Acad. Sci. USA* **90**, 11262–11266 (1993).
9. Higgs, D.R. *et al.* A major positive regulatory region located far upstream of the human α-globin gene locus. *Genes Dev.* **4**, 1588–1601 (1990).
10. Robertson, G. *et al.* Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci. USA* **92**, 5371–5375 (1995).
11. Hoess, R.H. & Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl. Acad. Sci. USA* **81**, 1026–1029 (1984).
12. Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl. Acad. Sci. USA* **93**, 5860–5865 (1996).
13. Palmiter, R.D. & Brinster, R.L. Germ-line transformation of mice. *Annu. Rev. Genet.* **20**, 465–499 (1986).
14. Araki, K., Araki, M., Miyazaki, J.-I. & Vassalli, P. Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **92**, 160–164 (1995).
15. Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. & Whitelaw, E. Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res.* **24**, 1465–1471 (1996).
16. Sutherland, H.G.E., Martin, D.I.K. & Whitelaw, E. A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene. *Mol. Cell. Biol.* **17**, 1607–1614 (1997).
17. Meyer, P., Heidmann, I. & Niedenhof, I. Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *Plant J.* **4**, 89–100 (1993).
18. Neuhuber, F., Park, Y.-D., Matzke, A.J.M. & Matzke, M.A. Susceptibility of transgene loci to homology-dependent gene silencing. *Mol. Gen. Genet.* **244**, 230–241 (1994).
19. Ye, F. & Signer, E.R. RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc. Natl. Acad. Sci. USA* **93**, 10881–10886 (1996).
20. de Carvalho, F. *et al.* Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11**, 2595–2602 (1992).
21. Ingelbrecht, I., Van Houdt, H., Van Montagu, M. & Depicker, A. Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl. Acad. Sci. USA* **91**, 10502–10506 (1994).
22. Bresnick, E.H. & Felsenfeld, G. Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA* **91**, 1314–1317 (1994).
23. Martin, D.I.K. & Whitelaw, E. The vagaries of variegating transgenes. *Bioessays* **18**, 919–923 (1996).
24. Grosveld, F., van Assendelft, G.B., Greaves, D.R. & Kollias, G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).
25. Ryan, T.M. *et al.* A single erythroid specific DNase I superhypersensitive site activates high levels of human β-globin gene expression in transgenic mice. *Genes Dev.* **3**, 314–323 (1989).
26. Matzke, M.A., Priming, M., Trnovsky, J. & Matzke, A.J.M. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649 (1989).
27. Rhounim, L., Rossignol, J.-L. & Faugeron, G. Epimutation of repeated genes in *Ascobulus immersus*. *EMBO J.* **11**, 4451–4457 (1992).
28. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.* **146**, 148–152 (1982).
29. Garrick, D., Sutherland, H., Robertson, G. & Whitelaw, E. Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Res.* **24**, 4902–4909 (1996).
30. Pondel, M.D., Proudfoot, N.J., Whitelaw, C. & Whitelaw, E. The developmental regulation of the human ζ-globin gene in transgenic mice employing β-galactosidase as a reporter gene. *Nucleic Acids Res.* **20**, 5655–5660 (1992).
31. Whitelaw, E. & Proudfoot, N. α-Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α2 globin gene. *EMBO J.* **5**, 2915–2922 (1986).

TAB 3

MOLECULAR AND CELLULAR BIOLOGY, Oct. 2000, p. 7419–7426 Vol. 20, No. 20
0270-7306/00/$04.00+0

Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells

SARA R. CHERRY,[1,2] D. BINISZKIEWICZ,[1] L. VAN PARIJS,[3] D. BALTIMORE,[3] AND R. JAENISCH[1,2]*

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142[1]; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[2]; and California Institute of Technology, Pasadena, California 91125[3]

Received 5 May 2000/Accepted 19 June 2000

Achieving long-term retroviral expression in primary cells has been problematic. De novo DNA methylation of infecting proviruses has been proposed as a major cause of this transcriptional repression. Here we report the development of a mouse stem cell virus (MSCV) long terminal repeat-based retroviral vector that is expressed in both embryonic stem (ES) cells and hematopoietic stem (HS) cells. Infected HS cells and their differentiated descendants maintained long-term and stable retroviral expression after serial adoptive transfers. In addition, retrovirally infected ES cells showed detectable expression level of the green fluorescent protein (GFP). Moreover, GFP expression of integrated proviruses was maintained after in vitro differentiation of infected ES cells. Long-term passage of infected ES cells resulted in methylation-mediated silencing, while short-term expression was methylation independent. Tissues of transgenic animals, which we derived from ES cells carrying the MSCV-based provirus, did not express GFP. However, treatment with the demethylating agent 5-azadeoxycytidine reactivated the silent provirus, demonstrating that DNA methylation is involved in the maintenance of retroviral repression. Our results indicate that retroviral expression in ES cells is repressed by methylation-dependent as well as methylation-independent mechanisms.

Retroviral vectors are appealing vehicles for gene transfer. However, long-term expression mediated by integrated proviruses in primary cells has been difficult to achieve. Retroviral regulatory elements are repressed in numerous cell types, including embryonic stem (ES) cells and hematopoietic stem (HS) cells (1, 3). For example, vectors that are functional in mature hematopoietic cells are often not expressed in blood cells of animals transplanted with the infected stem cells (18, 19, 31). In particular, the lack of significant provirus transcription in ES cells and their differentiated descendants has hampered the use of retroviral vectors in transgenic experiments (5, 12, 32). Interestingly, this block in provirus expression is maintained upon differentiation of infected cells despite the fact that primary infection of cells after differentiation results in efficient expression (6, 7, 26).

Transcriptional repression is thought to be mediated by both *cis*-acting de novo methylation of the integrated proviruses and cell-type-specific *trans*-acting transcriptional repressors (5, 9, 23). The effect of *trans*-acting factors on retroviral expression through binding of specific sequences within the promotors of retroviruses has been examined in many studies (29, 30, 35). In fact, the mouse stem cell virus (MSCV) long terminal repeat (LTR) was generated by the modification of the sequences within the LTR to increase the affinity for positive factors and decrease the affinity for negative regulators (20).

In contrast, the role of methylation in silencing has been less clear. DNA methylation is thought to be a general mechanism used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genomewide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in *Dnmt1* knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types.

ES cells provide a good model to study the role of DNA methylation in retroviral silencing. First, it was demonstrated that ES cells have high de novo methylation activity, which leads to effective methylation of integrated retroviral vectors, while little or no de novo methylation activity was detected in differentiated cells (21). In addition, ES cells were genetically modified to alter the endogenous level of DNA methylation by the targeted disruption of the maintenance methyltransferase gene *Dnmt1*. ES cells homozygous for this mutation proliferate normally with their genomic DNA highly demethylated, while differentiated cells and mice die due to the loss of genomic methylation (21, 22). Therefore, these modified ES cells are useful to study the effect of DNA methylation on retroviral gene expression. In addition, ES cells can be induced to differentiate in vitro or in vivo, allowing the study of DNA methylation and its effect on long-term expression.

Both Moloney virus-based and MSCV-based retroviral vectors have been used for gene transduction in a variety of cells. The MSCV vector is different from the typical Moloney virus vector in that the mutations in the LTR have allowed expression in a larger host range (8, 20). To this end, we modified MSCV to express the green fluorescent protein (GFP) as a sensitive reporter for gene expression (37). Using this vector, we demonstrated efficient expression in both ES and HS cells. We also demonstrated that silencing of retroviruses involves two mechanisms: (i) *trans*-acting factors that affect the initial expression of Moloney virus-based vectors but not MSCV-based vectors and (ii) long-term DNA methylation-dependent silencing that directly restricts expression of MSCV in ES cells and during embryogenesis. Silencing of the MSCV vector in wild-type ES cells and in in vivo differentiated ES cells was

* Corresponding author. Mailing address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Phone: (617) 258-5186. Fax: (617) 258-6505. E-mail: jaenisch@wi.mit.edu.

reversed by 5-azadeoxycytidine (5-azadC) treatments that demethylated the retroviral sequences, demonstrating that DNA methylation directly controls the maintenance of retroviral repression.

MATERIALS AND METHODS

Tissue culture. ES cells were cultured as described previously (21). To generate ES cell clones for injection into blastocysts, the ES cells were maintained on irradiated mouse embryonic fibroblasts (MEFs) with 500 U of leukemia inhibitory factor (LIF) per ml (22). For other experiments, the ES cells were cultured without MEFs in 1,000 U of LIF per ml. 293 cells were maintained in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum (FBS), penicillin, streptomycin, and glutamine. Abelson virus-transformed B cells were maintained in RPMI 1640 supplemented with 10% defined FBS (HyClone), penicillin, streptomycin, glutamine, and 50 μM β-mercaptoethanol. ES cells with retroviral integrants were in vitro differentiated as follows: the cells were passaged without LIF in the absence of MEFs on bacterial plastic petri dishes for 4 days, trypsinized, and cultivated with or without retinoic acid for 2 weeks (25).

Plasmids. The retroviral vectors MfgGFP, pMXGFP, and MSCViresGFP have been described elsewhere (27, 33, 37). The MSCViresGFP vector was modified by introducing either the Cre recombinase or the human Bcl-2 gene upstream of the internal ribosome entry site (IRES)-GFP cassette as described elsewhere (11, 37). The replication-incompetent helper plasmid pCL-eco was used (24).

Retroviral infections. To generate retroviral supernatants, 293 cells were transiently transfected by calcium phosphate-mediated coprecipitation with 5 μg of the replication-incompetent helper vector pCL-eco and 10 μg of the reporter retroviral vector as stated elsewhere (28). The cells were fed at 24 h postinfection, and the retroviral supernatant was used at 48 h. The cells continued to produce high-titer retroviruses for another 2 days, and that supernatant was used if needed for additional experiments. The supernatant was collected, brought to 4 μg of Polybrene per ml–10 mM HEPES, and filtered (0.45-μm-pore-size filter) for use.

ES cells for infection were washed and trypsinized. They were then plated at 10^6 cells per well of a six-well dish and centrifuged. The ES cell medium was removed, and retroviral supernatant was added at 1 ml/10^6 cells. Next, the plate was centrifuged for 45 min at 2,500 rpm at room temperature. The retroviral supernatants were removed; the cells were resuspended in ES cell medium and plated onto gelatinized dishes. ES cells used to generate mice were plated onto irradiated MEFs.

Bone marrow was infected as follows (36). C57BL/6 mice were obtained from the Jackson Laboratory (Bar Harbor, Maine). Bone marrow cells were harvested from the tibias and femurs of C57BL mice 5 days after they received an intraperitoneal injection of 5 mg of 5-fluorouracil (Sigma) in Dulbecco's phosphate-buffered saline (Gibco/BRL). These cells were then cultured for 4 days at 2×10^6 cells/ml with recombinant mouse interleukin-3 (rmIL-3; 20 ng/ml), rmIL-6 (50 ng/ml), and (50 ng/ml; recombinant mouse stem cell factor; R&D Systems) in Dulbecco's modified Eagle medium containing 10% FBS. After 48 and 72 h, the bone marrow cells were spin infected with the retroviral supernatant generated as described above. Then the retroviral supernatant was removed and replaced with growth medium containing cytokines.

FACS (fluorescence-activated cell sorting) analysis and sorting. Adherent cell lines were trypsinized, washed, and resuspended in complete medium to achieve a single-cell suspension at the time points indicated. Nonadherent cells were used directly for analysis. Organs were disrupted manually and passed through a 70-μm mesh to generate a single-cell suspension. The cells were analyzed for viability using scatter properties and the exclusion of propidium iodide. The level of GFP expression was monitored by fluorescence without compensation to detect cells with low levels of GFP expression. The ES cells were sorted into ES cell medium and plated immediately onto either gelatinized plates or MEFs for blastocyst injections. The survival of ES cells after sorting was approximately 50%, as measured by the number of colonies generated divided by the expected number of colonies.

5-AzadC treatments. ES cells were treated with 0.15 μM 5-azadC (Sigma) at days 1 and 3 postplating. The cells were fed, allowed to recover, and then assayed 4 to 8 days later. The red blood cells in whole blood were lysed (5), and the remaining cells were stained with fluorescently labeled anti-H2-b, anti-H2-d, anti-B220, anti-TCRa (Pharmingen) at 1:200 as indicated. At day 0, splenocytes were treated with either anti-CD3 or anti-CD40 (Pharmingen); 0.15 μM 5-azadC was added at day 1, and the anti-CD3-treated cells were assayed at day 4. 5-AzadC was added again to the B-cell cultures with fresh anti-CD40 at day 4, and the cells were assayed at day 6.

Staurosporine-mediated cell death. ES cells were infected with the stated retrovirus and treated with staurosporine at day 4 postinfection for 24 h with the indicated concentration of drug. The percentage of viable, GFP-positive cells was determined by flow cytometry (6). Data are presented as a percentage of GFP-positive cells before treatment. Results from one representative experiment of three performed are shown.

LacZ staining. ES cells were infected with the stated retrovirus and sorted for GFP expression at day 3 postinfection. The ES cells were plated and cultured for an additional 5 days and stained for LacZ expression as described elsewhere (41).

Adoptive transfers. Recipient mice (10) received a total of 1,200 rads of whole-body radiation in two doses (800 and 400 rads) 3 h apart and were then injected with 2×10^6 to 5×10^6 infected bone marrow cells. Irradiated mice were maintained on trimethaprim-sulfamethoxazole in sterile cages for 4 to 6 weeks to prevent opportunistic infections (34). Serial passages were performed by harvesting bone marrow from mice 6 to 8 weeks postreconstitution and transferring 2×10^6 to 5×10^6 cells into irradiated recipients. Mice were analyzed 8 to 12 weeks posttransfer to allow reconstitution of the T-cell compartment. These experiments were repeated multiple times with similar results.

Southern blot analysis. The genomic DNA was isolated as described elsewhere (19). Ten micrograms of DNA was digested with the stated restriction enzyme overnight. The products were resolved on an agarose gel, transferred to a nylon membrane, and detected using a probe that spans the entire GFP coding sequence.

RESULTS

High-efficiency retroviral expression in ES cells. Retroviral vectors based on the MSCV LTR were constructed with a multiple cloning site followed by an IRES driving expression of the gene for GFP as schematically diagrammed in Fig. 1A (MiG) (37). We generated high-titer retroviruses by transient transfection and infected ES cells with an adapted spin infection protocol. Using this protocol, we reproducibly achieved high-efficiency (>50%) infection of ES cells as measured by flow cytometry; uninfected control cells were negative for GFP expression (Fig. 1B). The intracellular concentration of GFP is directly proportional to the fluorescence intensity measured by flow cytometry.

Next, we compared expression of the MSCV-based retrovirus and Moloney virus-based retroviral vectors in ES cells. GFP expression was detectable with the MSCV LTR-containing MiG vector but not with the two Moloney virus-based viruses pMX (27) and Mfg (33) (Fig. 1B). This was not due to inefficient genomic integration of the provirus or to a lower titer. Southern blot analysis of genomic DNA demonstrated that all three proviruses were integrated in the ES cells (Fig. 1C). Also, when parallel B-cell cultures were infected with the retroviral supernatant used to infect ES cells, all of the retroviral vectors were expressed in B cells at comparable efficiencies (Fig. 1B).

GFP expression driven by the MSCV LTR in ES cells was substantially lower than in other differentiated cell lines tested (Fig. 1B and data not shown) (20). To determine whether this low level of expression was sufficient to drive functional expression of other gene products, we cloned the Cre recombinase upstream of the IRES-GFP cassette to generate MSCVCreiresGFP. We tested for Cre activity by infecting ES cells that contain a translational stop sequence flanked by *loxP* sites located between the Rosa 26 promoter and a *lacZ* reporter schematically diagrammed in Fig. 2A (34). If Cre is expressed at functional levels in these ES cells, the protein will catalyze recombination of the *loxP* sites, leading to loss of the stop sequences and the expression of LacZ. Indeed, we found that >99% of GFP-positive ES cells that were infected with the Cre-expressing retrovirus were also LacZ positive (Fig. 2B). Uninfected cells were both GFP negative and LacZ negative (data not shown). This indicates that virus-mediated gene transfer resulted in functional Cre expression.

Because Cre activity is required only transiently for LacZ expression, we tested a second gene product that must be stably expressed throughout the experiment. It has been demonstrated that Bcl-2 expression protects many cell types against staurosporine-mediated apoptosis (10). Therefore, we examined whether Bcl-2 could protect ES cells from apoptosis when expressed from the MSCV LTR. We cloned human Bcl-2 up-

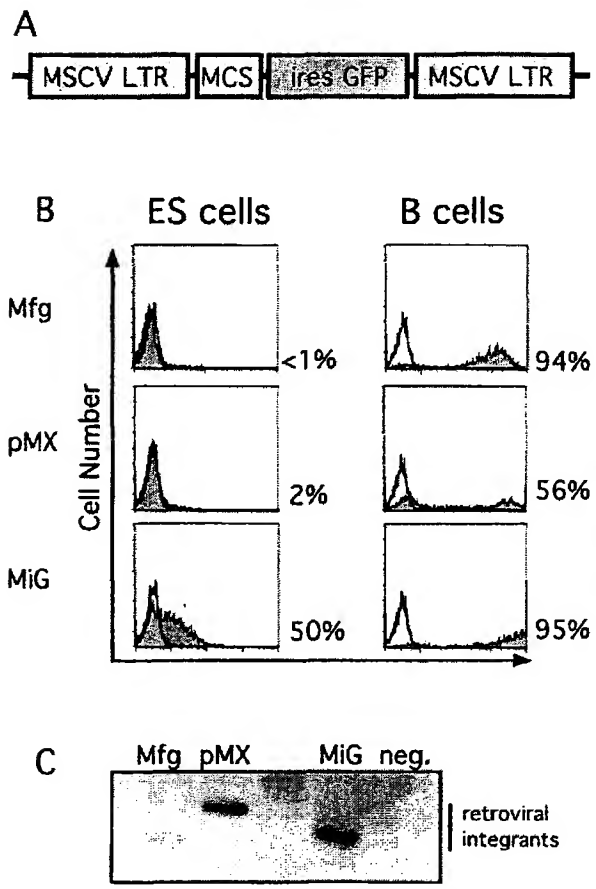

FIG. 1. Efficient retroviral infection of ES cells. (A) Schematic diagram of MiG vector containing the MSCV LTR followed by a multiple cloning site (MCS) and an IRES-GFP cassette. (B) MSCV-based (MiG) but not Moloney virus-based (Mfg and pMX) retroviruses express in ES cells. B cells or ES cells were infected by the indicated retroviruses and assayed by flow cytometry 2 days postinfection. Uninfected cells (unshaded) and infected cells (shaded area) were electronically gated for live cells and subsequently analyzed for GFP fluorescence and for cell number. Percentages of GFP-positive cells are indicated. (C) Comparable levels of integration of different retroviruses into ES cells, determined by Southern blot analysis of genomic DNA purified from infected and uninfected ES cells 2 days postinfection, digested with *Kpn*I, a restriction site present within the LTRs, and probed with the GFP coding sequence.

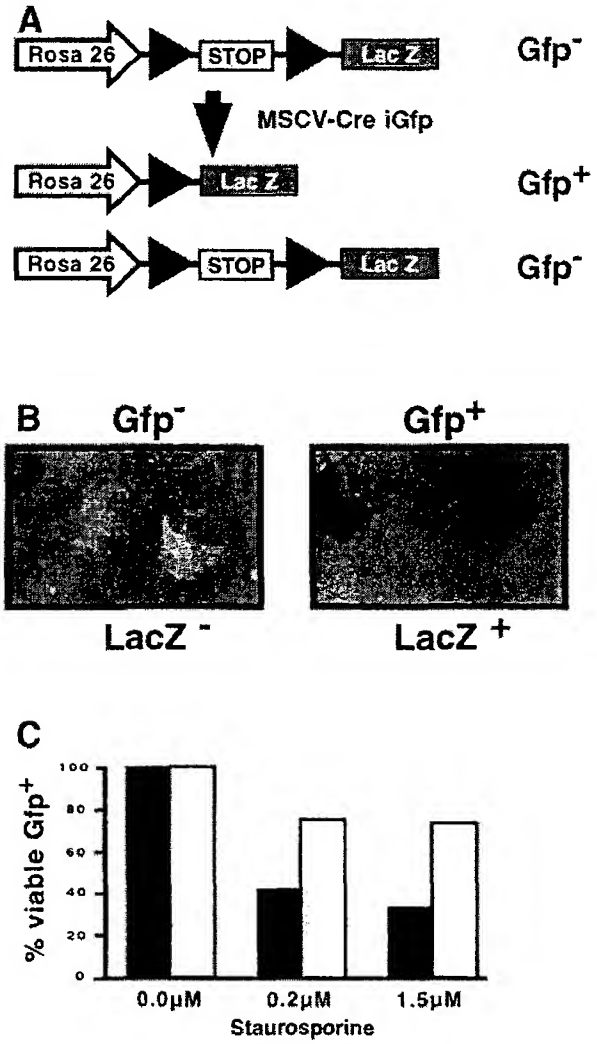

FIG. 2. Expression from the MSCV LTR is sufficient to drive functional gene expression. (A) Schematic diagram of the Rosa 26 locus in Cre reporter ES cells. Before Cre-mediated recombination, LacZ expression is prevented by the presence of a stop fragment. Retroviral infection with a Cre-expressing retrovirus with a GFP reporter results in two populations of cells. Cells that are GFP^+ become $LacZ^+$ due to efficient Cre-mediated recombination of the stop fragment. In contrast, cells that are GFP^- were not infected and thus remained $LacZ^-$. (B) ES cells were infected with the MSCVCreiresGFP retrovirus and sorted for either Gfp^- or Gfp^+ as indicated. The cells were subsequently cultured and stained for LacZ expression. Gfp^- cells are white (and therefore $LacZ^-$) while Gfp^+ cells are blue (and therefore $LacZ^+$). More than 99% of the Gfp^+ cells were $LacZ^+$ in multiple experiments. (C) ES cells were infected with either MSCViresGFP (■) or MSCVBcl-2iresGFP (□) and treated with the indicated amounts of staurosporine. The percentage of viable, Gfp^+ (infected) cells was determined by flow cytometry. The results are shown as a percentage of Gfp^+ cells before treatment. The results are from one representative experiment of three performed.

stream of the IRES-GFP cassette to generate MSCV Bcl-2iresGFP. Wild-type ES cells were infected with either the Bcl-2-expressing retrovirus or the control virus lacking Bcl-2. Increasing concentrations of staurosporine were added to the cultures, and flow cytometry was used to assay for both viability and GFP expression. GFP-positive cells infected with the Bcl-2-containing virus were significantly protected from staurosporine-mediated cell death compared to the GFP-negative cells or GFP-positive cells infected with the control retrovirus (Fig. 2C). Therefore, the level of expression from the MSCV LTR is sufficient for stable functional gene expression in ES cells.

Short-term transcriptional silencing in ES cells is methylation independent. It long has been hypothesized that retroviruses are transcriptionally silenced in embryonic cells by DNA methylation (12, 14, 21). Therefore, it was possible that DNA methylation of the MSCV LTR was responsible for the decreased level of expression in ES cells compared to other cell types (Fig. 1B). In addition, we sought to test whether DNA methylation of the Moloney virus-based vectors in the wild-type ES cells was the mechanism by which the Moloney virus-based LTRs were silenced (9, 13). To this end, we infected ES cells deficient for the maintenance DNA methyltransferase gene, *Dnmt1*, the loss of which results in genomewide hypomethylation (21, 22). *Dnmt1*$^{-/-}$ ES cells are demethylated, and proviral sequences remain unmethylated. The Moloney virus-based retroviruses such as pMX remained silent even when

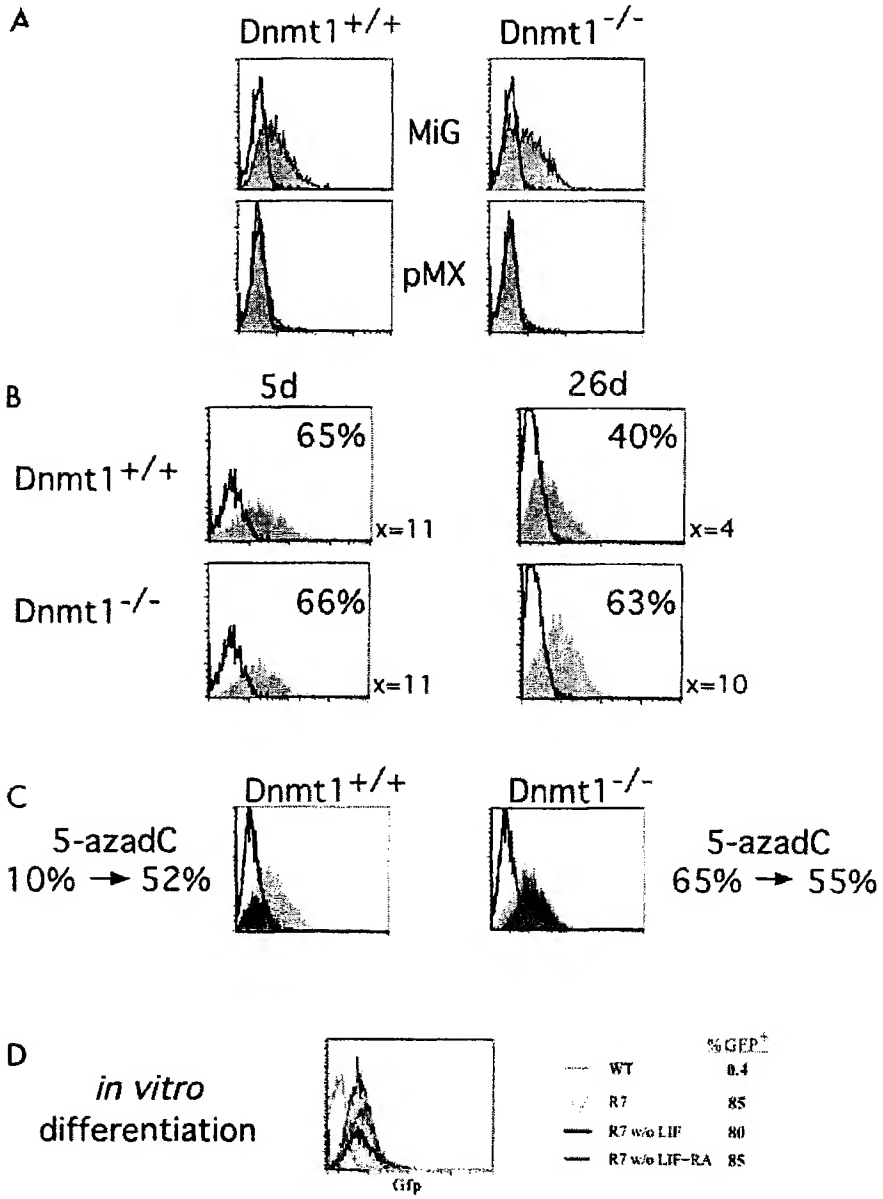

FIG. 3. Long-term expression of retroviruses is repressed by methylation. (A) MSCV-based (MiG) but not Moloney virus-based (pMX) retroviruses express in ES cells independent of the methylation status of the cells. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were not infected (unshaded) or infected by the indicated retroviruses (shaded) and assayed by flow cytometry 2 days postinfection as for Fig. 1B. (B) Long-term expression of GFP in ES cells is decreased in *Dnmt1*$^{+/+}$ cells but not *Dnmt1*$^{-/-}$ cells. The ES cells were infected with MiG, passaged for 5 or 26 days postinfection, and assayed by flow cytometry as above. The mean fluorescent intensity for the population and the percentage of GFP-positive cells are indicated. (C) Treatment with 5-azadC rescues the expression of retroviruses silenced by long-term passage. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were infected by MiG and passaged for >40 days. The cells were divided, and half were treated with 5-azadC. Then uninfected ES cells (unshaded), MiG-infected untreated ES cells (dark shading), and MiG-infected 5-azadC-treated ES cells (light shading) were assayed by FACS analysis. Numbers below the FACS plots are percentages of GFP-positive cells before and after 5-azadC treatment. (D) In vitro differentiation of ES cells does not affect retroviral expression. A clonal ES cell line (R7) infected with MiG or an uninfected wild-type (WT) ES cell control was in vitro differentiated by passage without feeders and LIF, with or without retinoic acid (RA) as indicated. The cells were assayed by flow cytometry, and the percentage of GFP-positive cells is indicated.

introduced into *Dnmt1*$^{-/-}$ ES cells, whereas MSCV expressed similar levels of GFP in both *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells (Fig. 3A). Therefore, the initial block in transcription directed by Moloney virus LTRs in ES cells is independent of DNA methylation and is presumably due to the binding of *trans*-acting factors. In addition, the mean fluorescence intensities of GFP were comparable between the *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells, indicating that the basal level of expression of the MSCV LTR is independent of DNA methylation.

DNA methylation constrains long-term retroviral expression. MiG-infected GFP-expressing ES cells were continually passaged to test the effect of DNA methylation on long-term expression.

Though GFP expression was high in both *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells at 5 days postinfection, a substantial fraction of the infected wild-type ES were GFP negative at 26 days postinfection. This was apparent by both a loss in the percentage of GFP-positive cells as well as a decrease in the mean fluorescence intensity of the bulk population of wild-type ES cells and was observed in both bulk cultures and individual cloned lines containing single integrants (Fig. 3B and data not shown). The fraction of GFP-positive cells continues to decrease with additional passages, as shown in Fig. 3C. These results suggest that long-term expression was suppressed by DNA methylation. To directly test whether retroviral repression was due to de novo methylation of the newly integrated retroviruses, we treated the long-term cultures with 5-azadC, a drug that leads to hypomethylation of genomic DNA (16). If DNA methylation was preventing expression of the MSCV LTR, treatment with the drug should activate retroviral expression. Indeed, we found that 5-azadC treatment of ES cells that had lost expression of GFP through long-term passage reactivated the provirus (Fig. 3C). In contrast, *Dnmt1*$^{-/-}$ ES cells infected with the retrovirus did not lose expression of GFP; thus, treatment with 5-azadC did not significantly affect retroviral expression (Fig. 3C). We also analyzed clonal lines containing single proviral integrants in which GFP expression was progressively silenced and found that treatment with 5-azadC resulted in the reactivation of gene expression in all cases (data not shown). This demonstrates that DNA methylation controls long-term but not short-term expression of retroviruses in ES cells.

Expression is maintained after in vitro differentiation. Previously, in vitro differentiation of ES cells had been demonstrated to silence expression of retroviral sequences (12, 20). Thus, we tested whether GFP expression from the MiG retrovirus in ES cells was affected by in vitro differentiation. We cultured MiG-infected wild-type ES cells in the absence of embryonic feeder cells and LIF in suspension to generate embryoid bodies. Disaggregated embryoid bodies were replated either with or without retinoic acid. We found no change in GFP expression in MiG-infected bulk cultures or individual subclones containing one to several integrants upon in vitro differentiation with either method, as shown for one clonal line containing multiple integrants in Fig. 3D. GFP expression was unchanged in all in vitro-differentiated ES cell lines, regardless of whether the subclones contained only a single or multiple integrants. This indicates that the MSCV-based MiG retrovirus is not silenced by in vitro differentiation.

Generation of mice from GFP-expressing MiG-infected ES cells. We next determined whether expression of the MSCV-based MiG vector was affected by in vivo differentiation of the infected ES cells. Cells from the chimeric animals were derived by injection of MiG-infected wild-type ES cells (derived from 129-Sv/Jae mice) into BALB/c blastocysts. MiG-infected *Dnmt1*$^{-/-}$ ES cells cannot be used for injection into blastocysts, because *Dnmt1*$^{-/-}$ ES cells die upon differention and therefore do not contribute significantly to adult mice (22). MiG-infected wild-type ES cells were sorted for GFP expression by flow cytometry prior to injection, and two GFP-expressing clones, R2 and R11, were isolated (Fig. 4B). Southern blot analysis demonstrated that R2 contained two integrants that comigrate on an agarose gel, and R11 contained three proviral integrants (Fig. 4A). High-contribution chimeras (>80% by coat color) were generated from the R2 and R11 ES cells, which transmitted the proviruses to their offspring (data not shown).

To test whether the chimeras expressed the integrated retroviruses, we isolated peripheral blood mononuclear cells

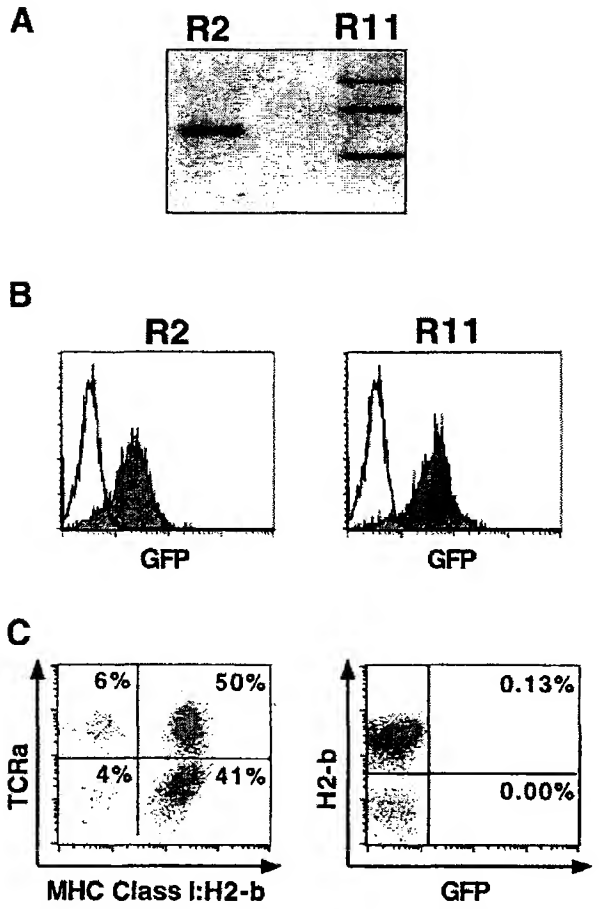

FIG. 4. Retrovirally infected, GFP-expressing ES cells generated nonexpressing mice. (A) Two retrovirally infected clones sorted for GFP expression were analyzed for proviral integrants by Southern blot analysis. R2 contained two integrants, while R11 contained three. Uninfected cells are negative. (B) The clones were passaged after sorting for GFP-expressing cells by flow cytometry and reanalyzed for GFP expression. Both R2 and R11 express GFP (shaded) compared to uninfected controls (unshaded). (C) PBMCs from the R2 chimera (more than 50% contribution by coat color) were analyzed by flow cytometry. ES cell contribution to the chimera was determined by phycoerythrin-H2-b staining and cyc-TCRa staining and demonstrated contribution to the T-cell compartment. The percentage of cells in each quadrant is listed. The cells were also monitored for GFP expression. The percentage of GFP$^+$ cells that are either major histocompatibility complex (MHC) class I H2-b$^+$ (ES cell derived) or H2-b$^-$ (blastocyst derived) is listed in the quadrant.

(PBMCs) from both the R2- and R11-derived chimeras. To distinguish whether the PBMCs were derived from the ES cell donor or the host blastocyst, we stained the cells with antibodies that recognized specific major histocompatibility complex class I haplotypes (Pharmingen). The donor ES cells (129 derived) are H2-b, and the blastocysts (BALB/C derived) are H2-d (Fig. 4C and data not shown). In addition, we stained the PBMCs with a pan-T-cell (TCRa) (Fig. 4C) or pan-B-cell (B220) antibody (data not shown) to determine the ES cell contribution to these lineages. Using this strategy, we found that approximately 90% of the PBMCs from either the R2 or R11 chimera were ES cell derived as measured by H2-b staining (Fig. 4C, data not shown). However, the majority of the cells did not express GFP in either chimera (Fig. 4C and data not shown). On the order of 0.1% of the PBMCs that were ES

cell derived were GFP positive, compared to less than 0.01% that were blastocyst derived (Fig. 4C). Similar results were also obtained with cells from the R11 chimera (data not shown). The results indicate that the MSCV LTR is repressed during in vivo differentiation to lymphocytes. Nevertheless, a small number of cells escaped silencing and expressed GFP. This transcriptional repression of the MiG provirus in the chimeras is in contrast to the GFP expression both in the donor ES and after in vitro differentiation (Fig. 4B, data not shown).

To determine if other somatic cells expressed the retroviral integrants, we analyzed the progeny of the chimeras. We isolated spleen, thymus, kidney, and liver cells from an animal carrying the two proviral integrants present in the R2 chimera and a littermate control containing no retroviral integrants. We analyzed these cells for GFP expression by FACS analysis and found no detectable expression of GFP in the splenocytes, thymocytes, renal cells, or hepatocyes (Fig. 5A and data not shown).

In vitro reactivation of retroviral expression. One possible explanation for transcriptional repression during in vivo differentiation was de novo methylation of the integrated retroviral LTR during embryonic development. To test this hypothesis, we cultured splenocytes from a mouse containing the R2 proviruses and from a littermate control, by treating the cells with either anti-CD3 or anti-CD40 to activate and induce proliferation of the T cells or B cells, respectively (4). We then assayed for GFP expression by flow cytometry and found that proliferation of the splenocytes did not activate expression of the retrovirus (data not shown). Next, we added 5-azadC to the splenocyte cultures to induce demethylation of the retroviruses. Indeed, treatment with 5-azadC activated expression in approximately 2% of the T cells (anti-CD3) (Fig. 5B) and 2% of the B cells (anti-CD40) (data not shown). In addition, when in vivo-differentiated cells, which had been isolated from the kidney of a transgenic mouse and transformed with simian virus 40 large T antigen (15), were treated with 5-azadC, activation of the silent provirus was observed in a similar fraction of the cells (data not shown). The extent of reactivation of expression of the provirus in in vivo-differentiated cells by 5-azadC was lower than in ES cells, where the reactivation of the provirus with 5-azadC was almost complete.

We next determined whether demethylation of the retrovirus in vivo would activate expression of the integrated retroviruses (13). Newborn mice were subcutaneously injected with 5-azadC at postnatal day 5 and subsequently analyzed at postnatal day 14 for GFP expression by flow cytometry. We found that 5-azadC-injected animals but not the uninjected controls had activated GFP expression of the proviruses in the spleen, thymus, and kidney (Fig. 5C and data not shown). When we injected higher concentration of 5-azadC in an effort to further demethylate the newborn mice, all injected animals died. This result demonstrated that repression by DNA methylation is, at least in part, responsible for silencing expression of the retroviral LTR in vivo.

Retroviral expression in HS cells after serial adoptive transfers. Bone marrow contains the HS cells that can stably repopulate the hematopoietic system after transfer to lethally irradiated mice. To determine whether HS cells can be effectively transduced and express the MiG retrovirus, we used infected bone marrow cells to reconstitute lethally irradiated mice (Fig. 6). We found that between 30 and 80% of the splenocytes from these primary recipients expressed the retrovirus, as measured by FACS analysis for GFP expression and shown for one representative experiment (Fig. 6A). The MiG virus was expressed in the B-cell, T-cell, and granulocyte compartments, as measured by a pan-B cell (B220), pan-T-cell (Thy-1), and pan-granulocyte (Gr-1) marker electronically gated on GFP-positive cells (Fig. 6B and data not shown). Because a large fraction of the splenocytes in the primary recipients are derived from relatively differentiated, lineage-committed progenitors, serial adoptive transfers are required to test for retroviral expression in the true HS cells (17). Therefore, we used bone marrow from these primary recipients to serially reconstitute lethally irradiated mice. This protocol requires substantial expansion from the stem cells and tests for long-term expression of the retrovirus. We observed no change in the percentage GFP-positive HS cells, and the level of GFP expression from the adoptive transfers into multiple recipients was stable over three additional passages (4° recipient). In addition, the infected cells gave rise to both B- and T-cell lineages at the expected ratios (Fig. 6B), demonstrating not only that the MiG retrovirus transduced the long-term repopulating HS cells but also that the MiG-mediated GFP expression was stable during in vivo hematopoietic differentiation. However, our results do not exclude the possibility that in addition to the transcriptionally active proviruses present within these cells, there are also copies of the virus that were transcriptionally silenced.

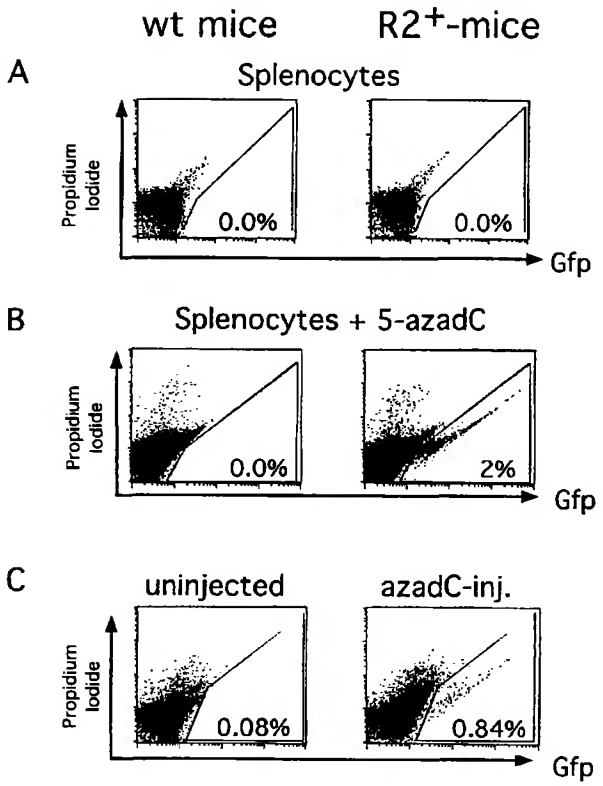

FIG. 5. Silenced retroviruses can be reactivated with 5-azadC. (A) Splenocytes from an $R2^+$ or $R2^-$ littermate do not express GFP by flow cytometry. Cells were stained with propidium iodide to exclude dead cells, and the percentage of GFP^+ cells is indicated. (B) The splenocytes from panel B were induced to proliferate with anti-CD3 and treated with 5-azadC. Cells were stained with propidium iodide to exclude dead cells and analyzed by flow cytometry. The percentage of GFP^+ cells is indicated. (C) Flow cytometric analysis of the splenocytes of littermates that were either uninjected or injected with 5-azadC at passage 5 and analyzed at passage 14 for GFP expression. The percentage of GFP^+ splenocytes is indicated.

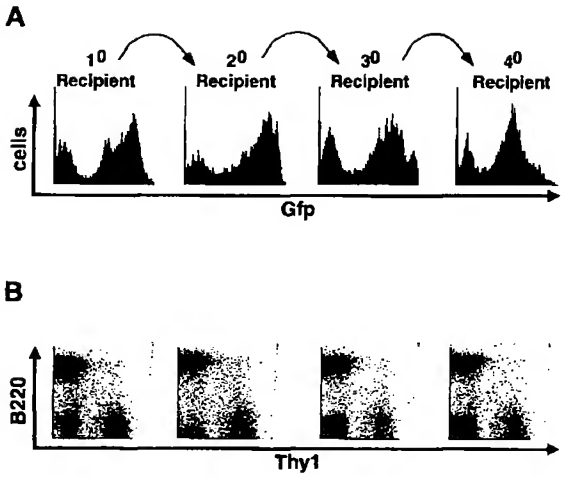

FIG. 6. Serial adoptive transfers maintain expression of the MSCV-based retrovirus. (A) Bone marrow was infected with MiG and used to reconstitute multiple lethally irradiated mice to generate the 1^0 recipient. The spleen of the 1^0 recipient was analyzed for GFP expression by flow cytometry. The bone marrow of the 1^0 recipient was used to reconstitute lethally irradiated 2^0 recipients. The spleen of a 2^0 recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 3^0 recipients. The spleen of a 3^0 recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 4^0 recipients. A representative analysis is shown. (B) splenocytes from panel A, stained with pan-B-cell (B220) and pan-T-cell (Thy-1) antibodies and electronically gated for GFP^+ cells, are shown below the GFP histogram they are derived from. The FACS diagrams are shown for these serially reconstituted spleens, demonstrating that the transferred cells contribute to both B- and T-cell lineages in the appropriate ratios.

DISCUSSION

We have investigated the role of DNA methylation in retroviral silencing. Retrovirus-based studies of stem cells have been hampered by the lack of expression. We have overcome the transcriptional repression in ES cells by using an MSCV-based vector in combination with a sensitive GFP reporter gene (MiG vector). The analysis of expression of the MiG vector and other Moloney virus-based vectors in *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells has allowed us to determine whether DNA methylation directly controls retroviral gene expression in these cells. We found that both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression.

Historically, retroviral expression of Moloney virus-based vectors in ES cells has been negligible. In contrast, the MSCV LTR not only transduces GFP expression in ES cells but also expresses other exogenous gene products such as the Cre recombinase and the antiapoptotic factor Bcl-2 at detectable level in ES cells. Therefore, the MSCV LTR can be used to express various transgenes in ES cells and their differentiated descendant cells.

It had been proposed that DNA methylation has evolved as a cellular mechanism to silence retroviral elements, preventing the spread of transposable elements through the genome (39). Indeed, de novo methylation of integrated proviral sequences has been observed in wild-type ES cells, which was correlated with the transcriptional silencing of the retrovirus (14). Our findings are the first demonstration that inhibition of the *Dnmt1* methyltransferase gene prevents silencing of the retroviruses in ES cells. This result provides direct evidence that DNA methylation is causally involved in long-term retroviral repression. Consistent with this conclusion is the demonstration that the transcriptionally silenced proviruses present in long-term *Dnmt1*$^{+/+}$ ES cell cultures can be reversed by treatments with 5-azadC.

In contrast, methylation-independent mechanisms determine initial retroviral expression in ES cells. Wild-type or *Dnmt1*$^{-/-}$ ES cells infected with Moloney virus-based vectors were transcriptionally silent, and therefore this silencing was independent of the DNA methylation status of the cells. Moreover, the basal level of expression from the MSCV-based vector was unaffected by the methylation status of the cells. This formally demonstrates that DNA methylation-independent mechanisms control initial retroviral gene expression in ES cells. Because the basal level of expression of the MSCV LTR in ES cells is lower than in differentiated cell types and not affected by the methylation status of the ES cells, *trans*-acting factors must regulate the initial level of expression.

Previous studies found that retroviruses, including the MSCV LTR, are silenced by the in vitro differentiation process (20). In contrast, we found for the first time that expression of this MSCV-based retrovirus in ES cells was maintained after in vitro differentiation with and without retinoic acid. We were also able to show long-term, stable GFP expression from the MiG vector in HS cells and their differentiated derivatives. MiG-mediated GFP expression from HS cells was stable through serial adoptive transfers, and the HS cells gave rise to GFP-expressing B- and T-cell lineages. Therefore, this MSCV-based retroviral transduction system should allow for a molecular analysis of stem cell biology and differentiation programs by forced expression of exogenous gene products.

It has been postulated that methylation-dependent mechanisms repress retroviral gene expression upon in vivo differentiation (13, 20). To test this, we injected GFP-expressing undifferentiated ES cells into recipient blastocysts and generated chimeric mice. Differentiated tissues derived from these in vivo-differentiated ES cells, such as PBMCs, lacked significant GFP expression. Treatment of ES cell-derived differentiated cells with 5-azadC in vitro or in vivo led to partial reactivation of expression of the silenced retroviruses in lymphoid and nonlymphoid tissues. We conclude from these results that the maintenance of retroviral silencing in vivo involves DNA methylation. However, only a small fraction of the 5-azadC-treated cells reactivated GFP expression, unlike the long-term ES cell cultures, in which every cell reactivated GFP expression. This suggests that methylation-independent mechanisms exist to suppress retroviral expression. Alternatively, 5-azadC treatment of differentiated cells, in contrast to ES cells, may not lead to a level of genomic demethylation sufficient for complete retroviral reactivation. The transgenic animals carrying the silenced MiG proviruses will be a valuable indicator for in vivo activation of GFP expression under different conditions.

ACKNOWLEDGMENTS

The first two authors, S. R. Cherry and D. Biniszkiewicz, contributed equally to this work.

We thank members of the Baltimore and Jaenisch labs for advice and discussions. We are grateful to Brian Bates for discussions and comments on the manuscript, George Daly for discussions, Jessie Dausman for blastocyst injections, Ruth Curry for mouse work, and Glen Paradis for FACS sorting.

D. Biniszkiewicz was supported by the Deutsche Akademische Austauschdienst. D. Baltimore and R. Jaenisch (NIH/NCI 5-R35-CA44339) are supported by NIH grants. This work was supported in part by the ERC program of the National Science Foundation under award EEC-9843342.

REFERENCES

1. Asche, W., G. Colletta, G. Warnecke, P. Nobis, S. Pennie, R. M. King, and W. Ostertag. 1984. Lack of retrovirus gene expression in somatic cell hybrids of

Friend cells and teratocarcinoma cells with a teratocarcinoma phenotype. Mol. Cell. Biol. 4:923–930.
2. **Cedar, H.** 1988. DNA methylation and gene activity. Cell 53:3–4.
3. **Challita, P. M., and D. B. Kohn.** 1994. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. Proc. Natl. Acad. Sci. USA 91:2567–2571.
4. **Coligan, J. E., et al. (ed.).** Current protocols in immunology. J. Wiley & Sons, New York, N.Y.
5. **Gautsch, J. W.** 1980. Embryonal carcinoma stem cells lack a function required for virus replication. Nature 285:110–112.
6. **Gautsch, J. W., and M. C. Wilson.** 1983. Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature 301:32–37.
7. **Greiser-Wilke, I., W. Ostertag, P. Goldfarb, A. Lang, M. Furusawa, and J. F. Conscience.** 1981. Inducibility of spleen focus-forming virus by BrdUrd is controlled by the differentiated state of the cell. Proc. Natl. Acad. Sci. USA 78:2995–2999.
8. **Grez, M., E. Akgun, F. Hilberg, and W. Ostertag.** 1990. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA 87:9202–9206.
9. **Hoeben, R. C., A. A. Migchielsen, R. C. van der Jagt, H. van Ormondt, and A. J. van der Eb.** 1991. Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblast cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. 65:904–912.
10. **Huang, D. C., S. Cory, and A. Strasser.** 1997. Bcl-2, Bcl-XL and adenovirus protein E1B19kD are functionally equivalent in their ability to inhibit cell death. Oncogene 14:405–414.
11. **Jacob, J., and D. Baltimore.** 1999. Modelling T-cell memory by genetic marking of memory T cells in vivo. Nature 399:593–597.
12. **Jaenisch, R., H. Fan, and B. Croker.** 1975. Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA 72:4008–4012.
13. **Jaenisch, R., A. Schnieke, and K. Harbers.** 1985. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. Proc. Natl. Acad. Sci. USA 82:1451–1455.
14. **Jahner, D., and R. Jaenisch.** 1985. Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. Nature 315:594–597.
15. **Jat, P. S., C. L. Cepko, R. C. Mulligan, and P. A. Sharp.** 1986. Recombinant retroviruses encoding simian virus 40 large T antigen and polyomavirus large and middle T antigens. Mol. Cell. Biol. 6:1204–1217.
16. **Jones, P. A.** 1985. Altering gene expression with 5-azacytidine. Cell 40:485–486.
17. **Jones, R. J., J. E. Wagner, P. Celano, M. S. Zicha, and S. J. Sharkis.** 1990. Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. Nature 347:188–189.
18. **Kohn, D. B.** 1995. The current status of gene therapy using hematopoietic stem cells. Curr. Opin. Pediatr. 7:56–63.
19. **Krall, W., and D. B. Kohn.** 1996. Expression levels by retroviral vectors based upon the N2 and the MFG backbones. Gene Ther. 3:365.
20. **Laker, C., J. Meyer, A. Schopen, J. Friel, C. Heberlein, W. Ostertag, and C. Stocking.** 1998. Host *cis*-mediated extinction of a retrovirus permissive for expression in embryonal stem cells during differentiation. J. Virol. 72:339–348.
21. **Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li.** 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195–3205.
22. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926.
23. **Loh, T. P., L. L. Sievert, and R. W. Scott.** 1990. Evidence for a stem cell-specific repressor of Moloney murine leukemia virus expression in embryonal carcinoma cells. Mol. Cell. Biol. 10:4045–4057.
24. **Naviaux, R. K., E. Costanzi, M. Haas, and I. M. Verma.** 1996. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. J. Virol. 70:5701–5705.
25. **Nichols, J., E. P. Evans, and A. G. Smith.** 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. Development 110:1341–1348.
26. **Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara.** 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
27. **Onishi, M., S. Kinoshita, Y. Morikawa, A. Shibuya, J. Phillips, L. L. Lanier, D. M. Gorman, G. P. Nolan, A. Miyajima, and T. Kitamura.** 1996. Applications of retrovirus-mediated expression cloning. Exp. Hematol. 24:324–329.
28. **Pear, W. S., G. P. Nolan, M. L. Scott, and D. Baltimore.** 1993. Production of high-titer helper-free retroviruses by transient transfection. Proc. Natl. Acad. Sci. USA 90:8392–8396.
29. **Petersen, R., G. Kempler, and E. Barklis.** 1991. A stem cell-specific silencer in the primer-binding site of a retrovirus. Mol. Cell. Biol. 11:1214–1221.
30. **Prince, V. E., and P. W. Rigby.** 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65:1803–1811.
31. **Robbins, P. B., D. C. Skelton, X. J. Yu, S. Halene, E. H. Leonard, and D. B. Kohn.** 1998. Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. Proc. Natl. Acad. Sci. USA 95:10182–10187.
32. **Robertson, E., A. Bradley, M. Kuehn, and M. Evans.** 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323:445–448.
33. **Silver, D. P., E. Spanopoulou, R. C. Mulligan, and D. Baltimore.** 1993. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. Proc. Natl. Acad. Sci. USA 90:6100–6104.
34. **Soriano, P.** 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. 21:70–71.
35. **Tsukiyama, T., H. Ueda, S. Hirose, and O. Niwa.** 1992. Embryonal long terminal repeat-binding protein is a murine homolog of FTZ-F1, a member of the steroid receptor superfamily. Mol. Cell. Biol. 12:1286–1291.
36. **Van Parijs, L., Y. Refaeli, A. K. Abbas, and D. Baltimore.** 1999. Autoimmunity as a consequence of retrovirus-mediated expression of C-FLIP in lymphocytes. Immunity 11:763–770.
37. **Van Parijs, L., Y. Refaeli, J. D. Lord, B. H. Nelson, A. K. Abbas, and D. Baltimore.** 1999. Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. Immunity 11:281–288.
38. **Walsh, C. P., J. R. Chaillet, and T. H. Bestor.** 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. 20:116–117.
39. **Yoder, J. A., C. P. Walsh, and T. H. Bestor.** 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13:335–340.

TAB 4

Gene, 91 (1990) 179–184
Elsevier

GENE 03596

The methylation-free status of a housekeeping transgene is lost at high copy number

(Recombinant DNA; hypomethylation; *HTF* island; hydroxy-methylglutaryl CoA reductase; cholesterol)

Majid Mehtali[a,b], Marianne LeMeur[a] and Richard Lathe[a,c]

[a] LGME-CNRS and U184-INSERM, 11 rue Humann, 67085 Strasbourg Cedex (France) Tel. 88-37 12 55; [b] Transgene SA, 11 rue de Molsheim, 67000 Strasbourg (France) Tel. 88 22 24 90; [c] AFRC Centre for Animal Genome Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541

Received by J.-P. Lecocq: 21 August 1989
Revised: 21 January 1990
Accepted: 22 February 1990

SUMMARY

Transgenic mouse lines were established bearing tandem arrays of a fusion construct comprising the promoter region of a housekeeping gene, *HMGCR*, encoding 3-hydroxy 3-methylglutaryl CoA reductase, linked to a bacterial *cat* reporter gene encoding chloramphenicol acetyltransferase (CAT). CAT activity was observed in all transgenic mouse tissues examined. The methylation state of the fusion transgene was determined. In non-transgenic mice the endogenous *HMGCR* promoter is devoid of methylation while flanking regions are extensively modified. In *HMGCR-cat* transgenic mice the fusion gene promoter was found to be similarly hypomethylated. However, the extent of hypomethylation varied with copy number: methylation-free status was progressively lost with increasing transgene copy number. Further transgenic mouse lines were constructed carrying a truncated *HMGCR* regulatory region linked to *cat*. Transgene expression and hypomethylation were observed in testis but not in any other tissue, and testis-specific methylation-free status was also lost at high copy number. Loss of hypomethylation at high copy number may indicate that saturable DNA-binding factors normally protect the *HMGCR* promoter from methylation.

INTRODUCTION

In contrast to the largely methylation-free genomes of bacteria and invertebrates, vertebrate genomes have extensive 5-methylation of C residues within CpG doublets (Grippo et al., 1968), and under-representation of CpG has been attributed to spontaneous deamination of 5-methyl-C (Bird, 1980). The relative insensitivity of vertebrate DNA to cleavage with enzymes whose recognition sequence contains a CpG doublet, for instance *Hpa*II (CCGG), is a consequence both of the rarity of CpG doublets and of cleavage inhibition by methylation of the recognition sequence. However, a small fraction of the genome is efficiently cleaved by *Hpa*II (Cooper et al., 1983) and such regions, '*Hpa*II-tiny-fragment' (HTF) islands, rich in unmethylated CpG doublets, appear to be associated with actively transcribed genes (Lavia et al., 1987) and particularly with the promoter regions of housekeeping genes (Bird, 1986; Gardiner-Garden and Frommer, 1987). Promoter regions of ubiquitously-expressed genes are unusual in that they are G/C-rich and lack consensus transcription initiation signals such as the 'TATA' or 'CAAT' boxes, and ubiquitous expression has been attributed to hypomethylation of control regions. (reviewed by Bird, 1986; 1987; Dynan, 1986; Gardiner-Garden and Frommer, 1987; Cedar, 1988). Methylation can correlate with diminished expression in vivo and in vitro (reviewed by Cedar, 1988);

Correspondence to: Dr. R. Lathe, CAGR, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541; Fax 31 667 0164.

Abbreviations: bp, base pair(s); CAT, Cm acetyltransferase; *cat*, gene encoding CAT; Cm, chloramphenicol; HMGCR, 3-hydroxy-3-methylglutaryl CoA reductase; *HMGCR*, gene encoding HMGCR; *HTF*, *Hpa*II tiny fragment; kb, kilobase(s) or 1000 bp; nt, nucleotide(s).

however, the causal link between hypomethylation and gene expression is unclear.

HMGCR (3-hydroxy-3-methylglutaryl CoA reductase) is a membrane-bound glycoprotein that catalyses a key step in the synthesis of cholesterol, an essential component of the cell membrane (Brown and Goldstein, 1980; Luskey, 1986). The *HMGCR* genes of hamster and human are characterized by a noncoding first exon and a long (ca. 3.5 kb) intervening sequence prior to the translation start site in the second exon (Reynolds et al., 1985; Luskey, 1987). We recently isolated the homologous mouse *HMGCR* gene and the promoter region, as for the hamster and human genes, bears the hallmarks of a *HTF* island in high G/C content and a CpG/GpC ratio close to unity (M.M. and R.L., unpublished).

We have used the transgenic mouse system to explore *HTF* island hypomethylation. Pronuclear microinjection of DNA into fertilized mouse eggs is now an established technique for generating mice bearing new gene combinations. In the majority of cases the injected DNA integrates at a single site as a multiple tandem repeat, and transgene copy number varies considerably between different transgenic lines (reviewed by Palmiter and Brinster, 1986). To explore possible variation in transgene functional status with copy number, two different fusion genes between the *HMGCR* promoter region and a *cat* reporter gene were constructed and introduced into the mouse germline. Employing these transgenic animals we endeavoured to study the relationship between transgene copy number, expression, and hypomethylation of the transgene promoter.

RESULTS AND DISCUSSION

(a) Transgenic animals bearing the *HMG1-cat* fusion construct

Construct *HMG1-cat* comprises a 5.5-kb *Bam*HI fragment of the mouse *HMGCR* promoter region, containing 1.35 kb of upstream sequence, the first (noncoding) exon and the first intron, linked to *cat* (Fig. 1A). The fusion construct was injected into fertilized mouse eggs and lines of transgenic mice were obtained carrying between 10 and 260 tandem copies of the transgene. These animals all express the fusion transgene in all tissues examined irrespective of transgenic line (Table I; also not presented): no proportionality was observed between expression level and copy number (M.M. and R.L., unpublished) as recorded (Palmiter and Brinster, 1986) in other transgenic systems.

(b) Methylation status of the *HMG1-cat* transgene

To assess the in vivo methylation status of the *HMGCR* promoter, we measured the extent to which template

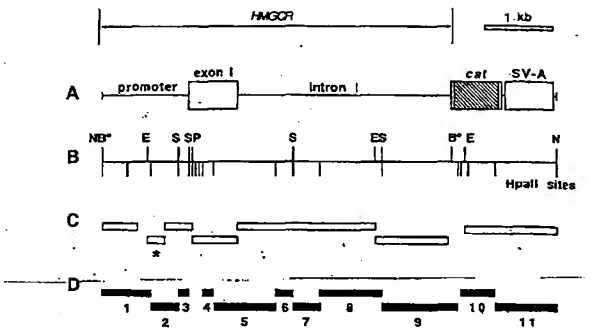

Fig. 1. Structure and analysis of the *HMGCR-cat* fusion transgene. (A) Structure of the injected construct *HMG1-cat*. The first *HMGCR* exon (noncoding) and the untranslated 5′ portion of the *HMGCR* second exon are indicated by open boxes; *cat*, hatched; the DNA sequence SV-A downstream from *cat* contains SV40 splice and polyadenylation signals derived from plasmid pSV2-CAT (Gorman et al., 1982). (B) Transgene restriction map. *Above*, restriction sites are B, *Bam*HI; B°, a *Bam*HI site destroyed during the cloning procedure; P, *Pvu*II, E, *Eco*RI, S, *Sma*I; flanking *Not*I sites (N) used for excision of the DNA construct prior to microinjection into fertilized mouse eggs are derived from the plasmid vector. *Below*, *Hpa*II/*Msp*I sites; the group of sites beneath the *Pvu*II site (P) is a cluster of seven sites over a region of 225 bp (spacing: 20, 25, 80, 30, 50 and 20 nt; unpublished data). (C) Hybridization probes employed to determine methylation status; the probe segment employed in the Southern analysis of Fig. 2 is marked with an asterisk. (D) DNA segments analysed for methylation inhibition of excision in Fig. 3.

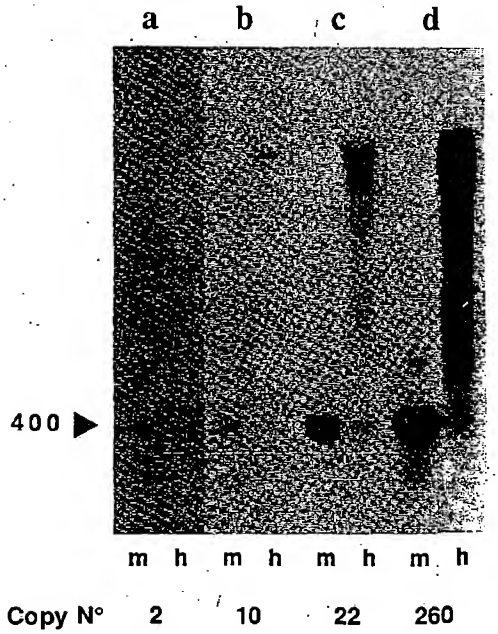

Fig. 2. Representative Southern analysis of *Hpa*II/*Msp*I digests. Liver DNA was prepared by a standard proteinase K/SDS/phenol procedure; 10 μg aliquots were digested with excess *Msp*I (m) or *Hpa*II (h), resolved by 1% agarose gel electrophoresis, and subjected to Southern analysis using a radiolabelled probe (marked with an asterisk in Fig. 1C) designed to detect the 400-bp segment 2 (arrowed) in Fig. 1D. Animals were: (a) wild-type; (b) transgenic line 74; (c) line 40; (d) line 78. The high-M_r bands of cleavage-resistant DNA are above the highest M_r marker employed (21.5 kb; not shown).

modification was able to inhibit excision of different transgene DNA segments from the genome. Total liver DNA from representative animals of different transgenic lines was cleaved either with *Hpa*II or with the methylation-insensitive isoschizomer *Msp*I, and after gel electrophoresis and Southern blotting examined for hybridization to separate probes (Fig. 1C) covering different regions of the *HMGCR* gene. The extent of methylation within each segment (Fig. 1D) of the transgene was assessed by the ratio of band intensities in the *Hpa*II and *Msp*I lanes. A typical experimental result is presented in Fig. 2 (zone analysed: segment 2 in Fig. 1D); data obtained for segments covering the entire transgene are compiled in Fig. 3.

Whereas excision of segments within the endogenous *HMGCR* promoter (two copies per diploid genome) in normal mice was not detectably blocked by methylation (<2%), flanking regions were resistant to digestion with *Hpa*II (Fig. 3a). In transgenic mice carrying ten copies of the *HMG1-cat* construct the *HMGCR* transgene promoter remained demethylated although the region devoid of methylation was narrower than that of the *HMGCR* gene in normal mice (Fig. 3b). However, methylation of the transgene promoter was observed to increase progressively in animals carrying 22 (35% methylation-inhibition of seg-

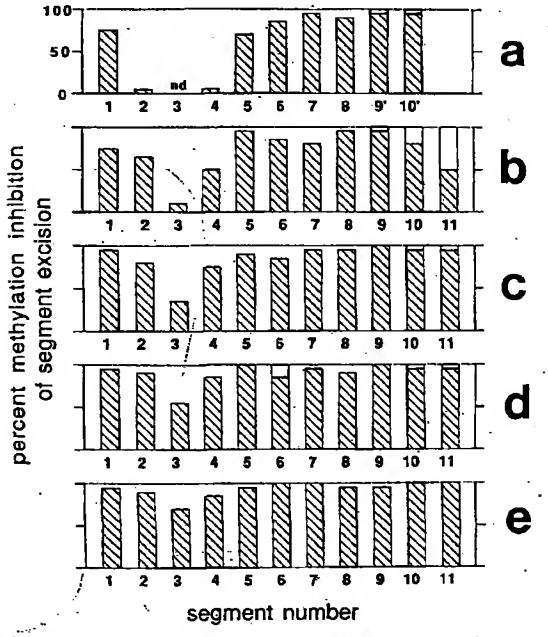

Fig. 3. Compilation of methylation data. Representative animals carrying different transgene copy numbers were: (a) wild type (2 copies of the *HMGCR* promoter); (b) transgenic line 74 carrying ten copies of the *HMG1-cat* fusion transgene); (c) line 40, 22 copies; (d) line 80, 45 copies; (e) line 78, 260 copies. Vertical axis: the extent of methylation inhibition of segment excision, determined as the ratio of the intensities of the relevant bands in the *Hpa*II and *Msp*I lanes of liver DNA (e.g, Fig. 2). Unshaded areas above certain histogram bars indicate 'greater-than' values. Horizontal axis: segment number (see Fig. 1D); internally-labelled probes used to determine methylation status (Mehtali, 1988) are presented in Fig. 1C. Quantitative scanning densitometry of autoradiograms (GS300 Scanner, Hoeffer Scientific) was used to determine the *Hpa*II/*Msp*I excision ratio. Segment 1 comprises two equally-sized *Hpa*II fragments that were not resolved by gel electrophoresis; segment 10 overlaps an *Hpa*II 'slow' site within *cat* that is partially refractory to *Hpa*II cleavage. Segments 9 and 10 represent the next two *Hpa*II fragments present in the endogenous *HMGCR* gene but not in the *HMG1-cat* transgene. For technical reasons the excision ratio for segment a3 was not determined (*nd*).

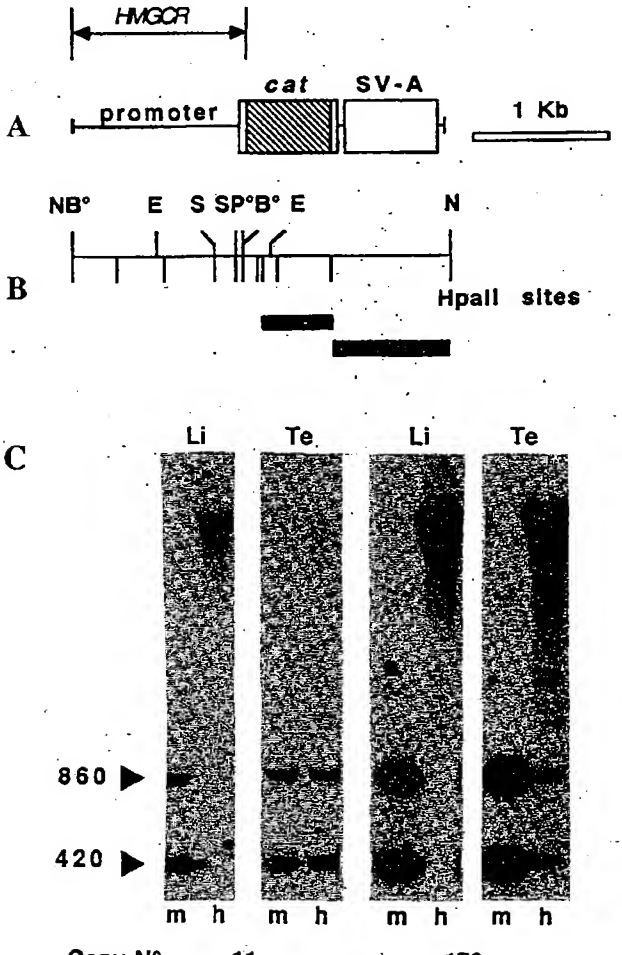

Fig. 4. Methylation state of the *HMG2-cat* transgene. (A) Structure of *HMG2-cat*: this construct is a derivative of *HMG1-cat*, in which the region between the *Pvu*II site in the *HMGCR* first exon and the site marked B° at the 5' end of *cat* in *HMG1-cat* (see Fig. 1A,B) has been removed. (B) Restriction site nomenclature is as in Fig. 1B; the solid bars given below indicate the segments detected by the hybridization probe employed (DNA fragment comprising *cat* and the SV40 splice and poly-adenylation signals, not shown). (Panels C) Southern hybridization of *Msp*I (m) and *Hpa*II (h) digests of total liver (Li) and testis (Te) DNA from representative animals bearing eleven (line 61) or 170 copies (line 76) of the fusion transgene. *Hpa*II fragments detected (sizes in bp) are arrowed. High-M_r band sizes and all experimental conditions are as in Fig. 2 legend.

TABLE I

CAT activity in tissues of *HMGCR-cat* transgenic mice

Construct[a]	Line	Copy No.[b]	CAT activity[d] in mouse tissues[c]						
			ta	li	in	ki	br	st	te
HMG1-cat	74[e]	10	12	5	21	5	62	31	59
HMG1-cat	40[e]	22	17	20	107	5	90	72	na
HMG1-cat	40	22	30	90	95	15	80	110	140
HMG1-cat	80[e]	45	6	32	21	7	70	23	na
HMG1-cat	78[e]	260	0.6	0.5	2	0.6	0.6	0.5	32
HMG2-cat	61[f]	11	0.6	0.9	0.7	0.6	0.7	0.8	135
HMG2-cat	76[f]	170	0.5	1.6	0.7	0.6	0.7	0.8	208

[a] Construct p*HMG1-cat* contains a 5.5-kb *Bam*HI segment of the mouse *HMGCR* gene (Mehtali, 1988) comprising the mouse *HMGCR* promoter region, the first (untranslated) exon, and the first intron, linked to the bacterial *cat* gene (Fig. 1A) and propagated in plasmid pPolyIII-i (Lathe et al., 1987). The downstream *Bam*HI site used in this construction lies immediately beyond the *HMGCR* exon II splice acceptor site (Mehtali, 1988; Gautier et al., 1989; M.M. and R.L., in preparation). In *pHMG2-cat* the region between the *Pvu*II site in the *HMGCR* first exon and the *BamHI*° site at the 5′ end of *cat* (see Fig. 1A) was removed: the structure of the fusion gene is presented in Fig. 4. Plasmid DNAs were propagated on *E. coli* 1106 (*thr leu thi hsdS*). *HMG-cat* fusion genes were excised from the vector by *Not*I digestion and purified by sucrose gradient centrifugation (10–30% sucrose, 35000 rpm, Beckman SW41 rotor, 14 h, 20°C). Two hundred copies were injected into fertilized eggs (C57B1/6 × SJLF1 hybrids) and the presence of the transgene was detected by Southern blot analysis of DNA from the tails of four-week-old animals (Palmiter et al., 1982). Transgenic lines were established by systematic back-crossing with C57B1/6 × SJLF1 hybrids.

[b] Copy numbers were determined by quantitative densitometry of Southern blots; all transgenic animals carry the transgene as an unrearranged tandem repeat integrated at a single autosomal location (data not presented).

[c] Tissues were: ta, tail; li, liver; in, intestine; ki, kidney; br, brain; st, stomach; te, testis.

[d] CAT activity in tissue homogenates (ultra-turrax, in 0.25 M Tris-HCl pH 7.8, 1 mM phenylmethylsulfonylfluoride) was determined, after heat treatment (10 min, 65°C) and clarification, by the transfer of radioactivity from ^{14}C acetyl CoA to ethyl-acetate-soluble Cm in a standard assay (Sleigh, 1986) at a constant protein concentration as measured using a commercial assay kit (BioRad), and is expressed as the incorporation of ^{14}C acetyl into Cm (cpm × 10^3) per 100 μg protein. The background value for CAT activity in negative extracts was 0.5. na, not applicable.

[e] Animals used for analysis in Fig. 2 (founder transgenic male of line 78; founder transgenic females of lines 40 and 80, and a second generation male for line 74; the remaining line 40 animal was a second generation male; M.M. and R.L., in preparation).

[f] Animals used for analysis were second generation males.

ment excision, Fig. 3c), 45 (48%) or 260 copies (78%) (Fig. 3d,e). The methylation pattern appeared identical in other tissues examined (intestine, testis) though the extent of the hypomethylated region was slightly larger in testis than in the other tissues (data not presented). The strength of the hybridization signal originating from the transgene unfortunately precluded examination of the methylation pattern of the endogenous *HMGCR* gene in these transgenic mice.

(c) *HMG2-cat* transgenic mice

To determine whether there is a general correlation between increased transgene copy number and loss of methylation-free status we constructed a deletion derivative of the *HMG1-cat* construct, *HMG2-cat*, in which a sub-fragment of the *HMGCR* promoter region is linked directly to *cat* (Fig. 4A). *HMG2-cat* was introduced into the mouse germline and tissues from transgenic animals were analysed for CAT activity (Table I). In the two lines examined (61 and 67; Table I) activity was detected in testis but not in other tissues. The methylation state of the fusion transgene in lines 61 and 76 was examined by *Hpa*II or *Msp*I digestion and Southern hybridization to probes designed to detect the reporter gene segment of the transgene (Fig. 4B) or the *HMGCR* promoter region (data not presented).

In liver (Fig. 4C) and intestine (data not presented), tissues in which no expression is observed (Table I), the fusion transgene was extensively methylated in both transgenic lines. In testis of mice harboring eleven copies of the transgene (line 61), a tissue in which the transgene is expressed, the *HMG2-cat* promoter and surrounding regions were essentially devoid of methylation (Fig. 4C, also data not presented). In contrast, the transgene in testis of line 76 (harboring 170 copies) was found to be substantially methylated (Fig. 4C), supporting a link between increased copy number and loss of methylation-free status. Nevertheless, in testis of line 76 a small proportion of transgene copies appeared to be devoid of methylation; these few unmethylated copies may be responsible for the observed transgene expression level. Indistinguishable methylation patterns were obtained in all cases irrespective of whether the probe employed covered the reporter gene segment (Fig. 4C) or the *HMGCR* promoter region (data not presented). It is of note that hypomethylation of the

HMGCR promoter region in testis of line 61 now extends into the adjacent bacterial *cat* gene (Fig. 4). Similar hypomethylation of *cat* was observed in testis of a further transgenic mouse line (No 60) harboring 15 copies of the *HMG2-cat* transgene (data not presented).

(d) Conclusions

HTF islands are associated with control regions of active genes, particularly the housekeeping genes (Bird, 1986; Lavia et al., 1987; Gardiner-Garden and Frommer, 1987). It was previously reported that the methylation-free *HTF* island of the *Thy-1* gene is maintained when the intact gene is introduced into the mouse germline (Kolsto et al., 1986). We describe here that hypomethylation of a housekeeping gene (*HMGCR*) promoter is dependent upon transgene copy number, and methylation-free status of the *HTF* island at the 5′ end of a *HMGCR-cat* fusion transgene, *HMG1-cat*, is lost at increasing copy-number. In further transgenic mice bearing the *HMG2-cat* deletion derivative of the fusion construct, CAT activity and transgene hypomethylation were only detected in testis. It is of note that ectopic expression of tissue-specific transgenes in testis has been observed previously (Lacy et al., 1983; Shani, 1986; Al-Shawi et al., 1988; and our unpublished data). As observed in *HMGCR-cat* transgenic animals, methylation-free status in testis of *HMG2-cat* transgenic animals was also lost at high copy number.

One possible explanation for loss of hypomethylation at high copy number is out-titration of regulatory proteins or other binding factors that protect the DNA from methylation. Our data do not exclude the possibility that out-titration of a demethylase activity (e.g., Razin et al., 1986) might also reduce methylation level. However, methylation of the transgene promoter at high copy number demonstrates that *HTF* island DNA is not intrinsically resistant, in vivo, to methylation of C residues within CpG doublets. Because saturable factors thus appear to protect DNA from methylation in vivo, factor binding seems likely to precede methylation. In consequence, it would appear unlikely that the methylation status of the *HMGCR* promoter itself plays a major role in determining the extent of factor binding in vivo.

Despite the fact that the *HMGCR* promoter present on the transgene can outnumber the endogenous *HMGCR* promoter by a factor of 100, we have not detected any alteration in the expression of the endogenous gene (M.M., R.L. and G. Boukamel, unpublished data), in agreement with the conclusions of a study (Davis and MacDonald, 1988) using a rat elastase I transgene. Because high transgene copy number appears to be without effect on the expression of the endogenous *HMGCR* gene, binding factors may be only locally out-titrated, possibly arguing for restricted diffusion of binding factors (see Richetti et al., 1988).

Although we report saturable hypomethylation of the *HMGCR* promoter, this result may contrast with the report of Kolsto et al. (1986) who describe hypomethylation of a hybrid *Thy-1* gene promoter in mice bearing 60 copies of the transgene. However, the presence of extraneous sequences in *Thy-1* transgenic mice (Grosveld and Kollias, 1988) complicates interpretation. It therefore remains unclear whether loss of methylation-free status and/or local out-titration of binding factors is a general feature of large transgene arrays.

We also report that the hypomethylated region of the *HMGCR-HTF* island can extend into adjacent bacterial DNA (*cat*). This phenomenon was only observed when the *cat* reporter gene was linked directly to the *HMGCR* promoter region (construct *HMG2-cat*) and not when the *cat* segment was separated from the *HMGCR* regulatory region by 4 kb of intervening *HMGCR* genomic DNA (construct *HMG1-cat*). Hypomethylation of adjacent bacterial DNA was only observed in testis, the only tissue in which transgene expression was detected, and we speculate that cooperative binding of factors (e.g., Phillips et al., 1989; see also Murray and Grosveld, 1987) to the *HMGCR* promoter and to adjacent CpG-rich bacterial DNA may be responsible. We cannot however exclude the possibility that transcription per se can contribute to under-methylation.

Taken together, our data argue that the cytosine-methylase passively methylates DNA according to its accessibility/affinity for the methylase. Because DNA methylation can inhibit gene activity, passive methylation of transcriptionally inactive regions may contribute to the repressed state. Lower eukaryotes and invertebrates lack detectable DNA methylation, and the large genome sizes of vertebrates and plants may provide a selective advantage for DNA methylation (Antequera and Bird, 1988), for instance by marking untranscribed and passively-methylated DNA for higher-order condensation.

ACKNOWLEDGEMENTS

We would like to thank C. Marfing for skillful microinjection and P. Meyer and I. Kuhn for technical assistance. C. Tomasetto, J.O. Bishop, A.J. Clark, A. Pavirani, P. Gerlinger and P. Chambon are thanked for helpful discussions. This work was supported by grants from the CNRS, INSERM, Centre de Lutte contre le Cancer, ARC and the FRMF.

REFERENCES

Al-Shawi, R., Burke, J., Jones, C.T., Simons, J.P. and Bishop, J.O.: A MUP promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. Mol. Cell. Biol. 8 (1988) 4821–4828.

Antequera, F. and Bird, A.P.: Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J. 7 (1988) 2295–2299.

Bird, A.P.: DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8 (1980) 1499–1504.

Bird, A.P.: CpG islands and the function of DNA methylation. Nature 321 (1986) 209–213.

Bird, A.P.: CpG islands as gene markers in the vertebrate nucleus. Trends Genet. 3 (1987) 342–347.

Brown, M.S. and Goldstein, J.L.: Multivalent feedback regulation of HMGCoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. J. Lipid Res. 21 (1980) 505–517.

Cedar, H.: DNA methylation and gene activity. Cell 53 (1988) 3–4.

Cooper, D.N., Taggart, M.H. and Bird, A.P.: Unmethylated domains in vertebrate DNA. Nucleic Acids Res. 11 (1983) 647–658.

Davis, B.P. and MacDonald, R.J.: Limited transcription of rat elastase I transgene repeats in transgenic mice. Genes Dev. 2 (1988) 13–22.

Dynan, W.S.: Promoters for housekeeping genes. Trends Genet. 2 (1986) 196–197.

Gardiner-Garden, M. and Frommer, M.: CpG islands in vertebrate genomes. J. Mol. Biol. 196 (1987) 261–282.

Gautier, C., Mehtali, M. and Lathe, R.: A ubiquitous mammalian expression vector, pHMG, based on a housekeeping gene promoter. Nucleic Acids Res. 17 (1989) 8389.

Gorman, C.M., Moffat, L.F. and Howard, B.H.: Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell. Biol. 2 (1982) 1044–1051.

Grippo, P., Daccararino, M., Parasie, E. and Scarano, E.: Methylation of DNA in developing sea urchin embryos. J. Mol. Biol. 36 (1968) 195–208.

Grosveld, F. and Kollias, G.: Ectopic expression of Thy-1 in the kidneys of transgenic mice induces functional and proliferative abnormalities. Cell 54 (1988) 920.

Kolsto, A.-B., Kollias, G., Giguere, V., Isobe, K.-I., Prydz, H. and Grosveld, F.: The maintenance of methylation-free islands in transgenic mice. Nucleic Acids Res. 14 (1986) 9667–9678.

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D. and Costantini, F.D.: Foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and in inappropriate tissues. Cell 34 (1983) 353–358.

Lathe, R., Vilotte, J.L. and Clark, A.J.: Plasmid and bacteriophage vectors for excision of intact inserts. Gene 57 (1987) 193–201.

Lavia, P., MacLeod, D. and Bird, A.: Coincident start sites for divergent transcripts at a randomly selected CpG rich island of mouse. EMBO J. 6 (1987) 2773–2779.

Luskey, K.L.: Structure and expression of 3-hydroxy 3-methylglutaryl coenzyme A reductase. Ann. N.Y. Acad. Sci. 478 (1986) 249–254.

Luskey, K.L.: Conservation of promoter sequence but not complex intron-splicing pattern in human and hamster genes for 3-hydroxy 3-methylglutaryl coenzyme A reductase. Mol. Cell. Biol. 7 (1987) 1881–1893.

Mehtali, M.: Ph.D. thesis, University of Strasbourg, 1988.

Murray, E.J. and Grosveld, F.: Site-specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation-mediated suppression. EMBO J. 6 (1987) 2329–2335.

Palmiter, R.D. and Brinster, R.L.: Germ line transformation of mice. Annu. Rev. Genet. 20 (1986) 465–499.

Palmiter, R.D., Chen, H.Y. and Brinster, R.L.: Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell 29 (1982) 701–710.

Phillips, S.E.V., Manfield, I., Parsons, I., Davidson, B.E., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., Saint-Girons, I. and Stockley, P.G.: Cooperative tandem binding of *met* repressor of *Escherichia coli*. Nature 341 (1989) 711–715.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D. and Cantoni, G.L.: Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA methylation during differentiation. Proc. Natl. Acad. Sci. USA 83 (1986) 2827–2831.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G. and Luskey, K.L.: HMG CoA reductase: a negatively-regulated gene with unusual promoter and 5′ untranslated regions. Mol. Cell. Biol. 7 (1985) 1881–1893.

Richetti, M., Metzger, W. and Heumann, H.: One-dimensional diffusion of *Escherichia coli* DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. Proc. Natl. Acad. Sci. USA 85 (1988) 4610–4614.

Shani, M.: Tissue-specific and developmentally-regulated expression of a chimeric actin-globin gene in transgenic mice. Mol. Cell. Biol. 6 (1986) 2624–2631.

Sleigh, M.: A non-chromatographic assay for expression of the chloramphenicol-acetyl-transferase gene in eucaryotic cells. Anal. Biochem. 156 (1986) 257–262.

TAB 5

Cell, Vol. 32, 1105–1113, April 1983, Copyright © 1983 by MIT 0092-8674/83/041105-09 $02.00/0

Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells

Ohtsura Niwa* Yoshifumi Yokota,*
Hiroyuki Ishida* and Tsutomu Sugahara†
* Department of Experimental Radiology
Faculty of Medicine
Kyoto University
Sakyo-ku, Kyoto 606, Japan
† Kyoto National Hospital
Fushimi-ku, Kyoto 612, Japan

Summary

Expression and DNA methylation of the Moloney murine leukemia virus (M-MuLV) genome were investigated in murine teratocarcinoma cells after virus infection. The newly acquired viral genome was devoid of methylation, yet its expression was repressed. The integrated viral genome in undifferentiated teratocarcinoma cells was methylated within 15 days after infection. Although 5-azacytidine decreased the level of DNA methylation, it did not activate M-MuLV in undifferentiated cells. Activation by 5-azacytidine occurred only in differentiated teratocarcinoma cells. Thus two independent mechanisms seem to regulate gene expression during the course of differentiation. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogeneously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated.

Introduction

Laboratory strains of mice as well as field isolates of Mus species carry a large number of endogenous type C virus genomes. Blot hybridization analysis of mouse DNA digested with various restriction enzymes suggested that cells of inbred mice have as many as 50 copies of DNA sequences related to type C viruses (Dolberg et al., 1981). Inducible murine endogenous viruses are classified into two host range groups, ecotropic viruses and xenotropic viruses. Chromosomal locations of ecotropic viruses can be determined by classical Mendelian genetics. Ecotropic virus loci of AKR mice, *Akv1* and *Akv2*, have been mapped on chromosome 7 (Rowe et al., 1972) and on chromosome 16 (Kozak and Rowe, 1980), respectively. The *Cv* locus, an ecotropic virus locus of C3H/HeJ and BALB/c strains was detected on chromosome 5 (Kozak and Rowe, 1979; Ihle et al., 1979). C3H/Fg mice have an ecotropic virus locus on chromosome 7 at a site distinct from the *Akv1* locus. C58 and B10·BR mice carry unnamed ecotropic virus loci on chromosome 8 and on chromosome 11, respectively. Thus the distribution of endogenous ecotropic virus loci over a wide range of mouse chromosomes suggests that these endogenous virus genomes may have recently derived from a common prototype virus through germline integration during establishment of laboratory mice. Actually, Jaenisch demonstrated that infection of preimplantation stage BALB/129 mouse embryos with M-MuLV leads to integration of the virus into the germ line, in an endogenous unexpressed state (Jaenisch, 1976).

Infection of mouse embryo fibroblasts with murine ecotropic viruses usually results in stable integration of the viral genome and subsequent production of progeny virus. This makes for a clear distinction between fibroblast lines and the cells of preimplantation stage embryos where infection with exogenous viruses results in repression of the integrated virus genomes. Murine teratocarcinoma cells are the malignant and pluripotent stem cells derived from carcinoma of preimplantation stage embryos (Lehman et al., 1974). Teratocarcinoma stem cells infected with ecotropic murine type C viruses do not produce progeny virus (Peries et al., 1977; Teich et al., 1977; Huebner et al., 1979; Gautsch, 1980). The infecting viral genome, however, is carried in these cells and can be activated with bromodeoxyuridine (BrdUrd) (Speers et al., 1980).

The expression of endogenous virus genomes in avian and mouse embryo fibroblasts is regulated by DNA modification and can be activated by treating the cells with 5-azaCytidine (5-AzaCyd), a potent inhibitor of DNA methylation (Groudine et al., 1981; Niwa and Sugahara, 1981). We studied the state of the M-MuLV genome in teratocarcinoma cells and found that the viral genome in undifferentiated stem cells is repressed not by DNA methylation but by some other mechanism. Only after differentiation of the cells is the viral sequence under the control of a mechanism regulated by DNA modification.

Results

Repression of M-MuLV in Undifferentiated Teratocarcinoma Cells, and Induction of Its Expression by BrdUrd and Retinoic Acid

EC-A1 cells derived from PCC4 cells, a pluripotent stem cell line of mouse teratocarcinoma cells, were infected with M-MuLV. Cells were treated for 24 hr with BrdUrd and/or retinoic acid, a compound that induces differentiation of teratocarcinoma cells (Strickland and Mahdavi, 1978) before or after virus infection. Since both BrdUrd and retinoic acid exerted potent cytotoxic effects on EC-A1 cells, the virus-producing cells in the drug-treated cultures may be preferentially eliminated to yield a false negative result. Thus EC-A1 cells were treated with BrdUrd and/or retinoic acid before or after virus infection, and the cells were then cocultivated with SC-1 cells. The SC-1 cells were passaged once and assayed for virus

expression. The results in Table 1 indicate that preinfection and postinfection treatment with BrdUrd alone was effective for productive infection of the virus. Preinfection treatment with retinoic acid made the EC-A1 cells susceptible to infection with M-MuLV. The expression of the virus in the BrdUrd-treated cells was enhanced by the presence of retinoic acid. Although less efficient, postinfection treatment of the cells with BrdUrd together with retinoic acid did induce expression of the virus. A low but definite production of M-MuLV was detected by amplification through SC-1 cells when the EC-A1 cells were treated for 24 hr with 0.3–1 M of retinoic acid alone, immediately after infection.

Infection of F-9 cells, a nullipotent teratocarcinoma cell line (Bernstine et al., 1973), with M-MuLV also required preinfection or postinfection treatment with BrdUrd for efficient expression of the virus (O. Niwa, unpublished observation).

The teratocarcinoma stem cells, EC-A1 and F-9 cells, were all derived from 129 mice that contained the ecotropic virus sequences and were mostly subgenomic in size (Chan et al., 1980). These have been classified as a no virus strain (Chattopadhyay et al., 1974). The virus which was recovered from BrdUrd-treated EC-A1 cells, and which had been infected with M-MuLV, grew equally well on NIH/3T3 and BALB/3T3 cells. The tropism of the virus and the lack of virus activation from uninfected EC-A1 cells (O. Niwa, unpublished observation) strongly suggest that the virus thus recovered from infected EC-A1 cells after BrdUrd treatment is M-MuLV and not the virus endogenous to EC-A1 cells.

Isolation of EC-A1(Mo) Clones Carrying M-MuLV Genome in a Repressed State

EC-A1 cells were infected with M-MuLV at a moi of 1.0, which had been determined by titration on SC-1 cells. The cultures were then trypsinized, and the cells were plated for ring cloning. Each of the randomly isolated clones was treated with 20 μg/ml of BrdUrd and cocultivated with SC-1 cells. SC-1 cells were passaged three times and tested for virus expression by the reverse XC test. Of 110 clones thus tested, 47 expressed virus after BrdUrd treatment. These virus-inducible clones, designated EC-A1(Mo) clones, therefore carry the M-MuLV genome, in a repressed form.

Transcriptional Control of M-MuLV Expression in Undifferentiated Teratocarcinoma Cells

Liquid hybridization experiments indicated no detectable viral RNA transcript in M-MuLV-infected teratocarcinoma cells (Teich et al., 1977). We have isolated total cellular RNA from one of the virus-carrying clones, EC-A1(Mo)4. RNA was size-differentiated on agarose gel electrophoresis, transferred to diazobenzyloxymethyl paper (DBM paper) and analyzed for the sequence hybridizable to ^{32}P-labeled M-MuLV DNA (Figure 1). SC-1 cells productively infected with M-MuLV had two major bands, 34S and 24S, corresponding to virus genomic RNA and spliced *env* gene messenger RNA. On the other hand, uninfected SC-1 cells, EC-A1 cells and EC-A1(Mo)4 cells contained no detectable level of RNA sequence hybridizable to M-MuLV DNA. The same filter was hybridized with an 18S rRNA probe. Three bands corresponding to 18S rRNA and two precursor RNAs were detected in all four RNA samples. The density of the bands revealed by this probe varied little among the four cell lines tested, indicating that the amounts of RNA used for the analysis were the same for these cell lines. Total cellular RNA isolated from EC-A1 cells 48 hr after infection with M-MuLV was analyzed by the dot blot hybridization technique (data not shown). Here too, the RNA transcript of M-MuLV was not detected.

Table 1. Plaque-Forming Cells per 2×10^5 SC-1 Cells Cocultivated with EC-A1 Cells Treated Preinfection and Postinfection with BrdUrd and/or Retinoic Acid

BrdUrd (μg/ml)	Retinoic Acid (μM)	Preinfection Treatment/ Postinfection Treatment
0	0	0/0
0	0.03	5/0
0	0.1	14/0
0	0.3	23/5
0	1.0	46/3
5	0	122/7
5	0.03	634/7
5	0.1	TMTC/25
5	0.3	TMTC/31
5	1.0	TMTC/23
10	0	TMTC/6
10	0.03	TMTC/15
10	0.1	TMTC/43
10	0.3	TMTC/22
10	1.0	TMTC/19
20	0	TMTC/12
20	0.03	TMTC/21
20	0.1	TMTC/48
20	0.3	TMTC/39
20	1.0	TMTC/51

TMTC: too numerous to count.

5-AzaCyd Induction of M-MuLV Gene Expression and the State of Differentiation of Teratocarcinoma Cells

We attempted to induce M-MuLV expression in undifferentiated EC-A1(Mo) clones by 5-AzaCyd. To our surprise, none of the 47 EC-A1(Mo) clones expressed virus after treatment with 2 μg/ml 5-AzaCyd, although virus was readily recovered when the clones were treated with BrdUrd.

To determine whether or not the DNA methylation was suppressed by 5-AzaCyd in undifferentiated teratocarcinoma cells, two clones, EC-A1(Mo)1 and EC-A1(Mo)4, were treated with 5-AzaCyd, and the level of DNA methylation was measured. As is clear from

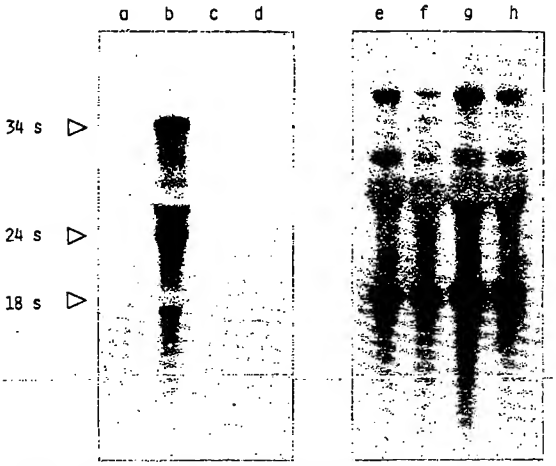

Figure 1. Blot Hybridization of Total Cellular RNA Isolated from Four Cell Lines

Probes used were ^{32}P-labeled M-MuLV DNA (lanes a, b, c and d) and ^{32}P-labeled cloned rRNA gene (lanes e, f, g and h). (Lanes a and e) RNA from SC-1 cells; (lanes b and f) RNA from SC-1 cells infected with M-MuLV; (lanes c and g) RNA from EC-A1 cells; (lanes d and h) RNA from EC-A1(Mo)4 cells.

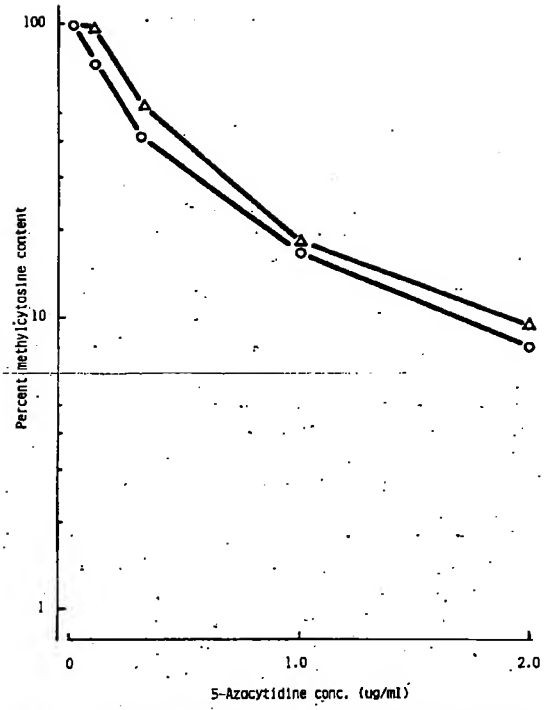

Figure 2. Level of Methylcytosine in Newly Replicated DNA of EC-A1 Cells

Cells were seeded onto 3 cm dishes at a concentration of 1 × 10^5 cells per dish. After overnight incubation, the dishes were nourished with medium containing (methyl-^{3}H)methionine at 4 μCi/ml and (2-^{14}C)thymidine at 0.005 μCi/ml. DNA was collected, hydrolyzed and separated on a cellulose thin-layer glass plate. The ratio of ^{3}H and ^{14}C counts in methylcytosine and thymidine respectively was taken as a relative measure of methylcytosine content (○) EC-A1(Mo)1 cells; (△) EC-A1(Mo)4 cells.

Figure 2, the level of methylcytosine in cellular DNA decreased with increasing doses of 5-AzaCyd. Therefore, in undifferentiated teratocarcinoma cells, suppression of DNA methylation does not seem to result in activation of the viral genome.

EC-A1(Mo)4 cells were treated with dimethylacetamide for 2 weeks to induce differentiation, and such was assessed to be complete by the morphology. M-MuLV was never expressed in the differentiated EC-A1(Mo)4 cells. These cells were then treated with 5-AzaCyd and cocultivated with SC-1 cells. The SC-1 cells were passaged twice, and the reverse XC test was performed at each passage. Only in the differentiated EC-A1(Mo)4 cells was the 5-AzaCyd-activated expression of M-MuLV evident (Table 2). Therefore, at least in the differentiated EC-A1(Mo)4 cells, the viral genome seems to be regulated by the extent of DNA methylation, and suppression of DNA methylation activates expression of the virus.

Table 2. Plaque-Forming Cells per 2 × 10^5 SC-1 Cells Cocultivated with EC-A1(Mo)4 Cells Treated with 5-AzaCyd

Cells	Concentration of 5-AzaCyd (μg/ml)	First Passage/ Second Passage
EC-A1(Mo)4 cells undifferentiated	2	0/0
	4	0/0
	6	0/0
	8	0/0
EC-A1(Mo)4 cells differentiated	1	0/0
	2	2/31
	4	13/TMTC

TMTC: too numerous to count.

Lack of DNA Methylation of Unintegrated M-MuLV Genome

The data presented above indicate that the M-MuLV genome in the differentiated teratocarcinoma cells is methylated. Experiments were then designed to determine the timing of DNA methylation of the M-MuLV genome, after virus infection of the undifferentiated cells.

EC-A1 cells were infected with M-MuLV, and DNA was extracted from the Hirt supernatant fraction which contained unintegrated viral genome. DNA isolated from EC-A1 cells 6 hr after infection contained three molecular species of M-MuLV—namely, closed circular, linear and open circular DNAs (Figure 3, lane a). Closed circular DNA and open circular DNA consisted of two subbands differing slightly in size. A linear molecule had a single band of 8.8 kb. Upon cleavage with Hind III, two bands corresponding to 8.2 kb and 8.8 kb linear molecules were detected (Figure 3, lane b) and were assumed to be the full-sized M-MuLV genome with one and two long terminal repeats (LTRs). All of these sequences were com-

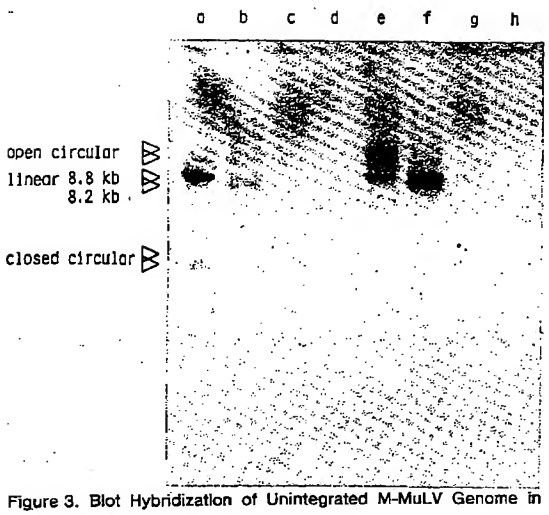

Figure 3. Blot Hybridization of Unintegrated M-MuLV Genome in Undifferentiated EC-A1 Cells

EC-A1 cells were infected with M-MuLV. DNA was extracted from the Hirt supernatant fraction of EC-A1 cells 6 hr after infection with M-MuLV (lanes a, b, c and d), or 12 hr after infection (lanes e, f, g and h). The probe used was ^{32}P-labeled M-MuLV DNA. (Lanes a and e) undigested DNA; (lanes b and f) digestion with Hind III; (lanes c and g) digestion with Hap II; (lanes d and f) digestion with Msp I.

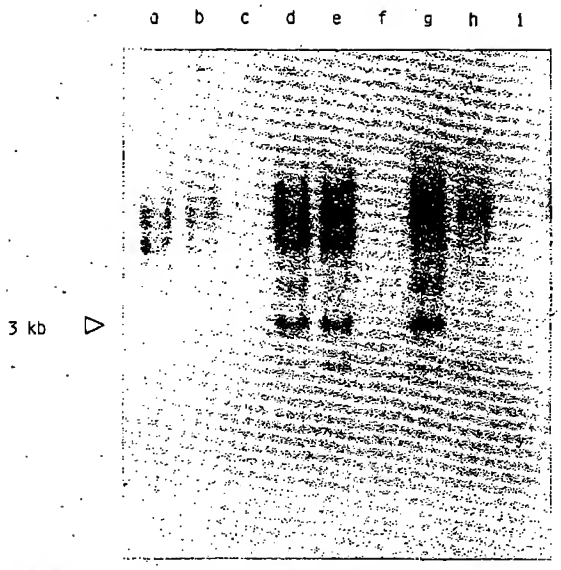

Figure 4. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from uninfected EC-A1 cells (lane a, b and c), EC-A1(Mo)4 cells (lane d, e and f) or EC-A1 cells 48 hr after infection with M-MuLV (lanes g, h and i). The probe used was the ^{32}P-labeled Sma I fragment of M-MuLV DNA. (Lanes a, d and g) digestion with Bam HI; (lanes b, e and h) digestion with Bam HI plus Hap II; (lanes c, f and i) digestion with Bam HI plus Msp I.

pletely digested by Hpa II and Msp I, suggesting that they are devoid of methylation at the CCGG sequence (Figure 3, lanes c, d). Similar results were obtained with DNA isolated at 12 hr after infection (Figure 3, lanes e, f, g, h). However, one difference in the 12 hr sample was that this DNA lacked a closed circular molecule (Figure 3, lane e). It is of interest that the Hirt supernatant DNA isolated 24 hr after infection had a much lesser amount of viral DNA, while the 48 hr sample was devoid of M-MuLV sequences (data not shown). In the case of SC-1 cells, DNA extracted 24 hr after infection had the greatest amount of unintegrated M-MuLV provirus sequence (O. Niwa, unpublished observation).

Undermethylation of Freshly Integrated M-MuLV Genome and Its Subsequent Methylation during Multiple Cell Cycling of Undifferentiated Cells

EC-Al cells were infected with M-MuLV at a moi of 2, and DNA was isolated 48 hr later. The integration of viral genome occurs at random sites. Digestion of DNA from randomly infected cells with Eco RI, which does not cleave the M-MuLV genome, will produce multiple fragments of various sizes carrying integrated viral sequences flanked by cellular sequences, and these cannot be resolved by agarose gel electrophoresis. Therefore, we used Bam HI, which cut the M-MuLV genome internally. As a probe to detect viral genome, the 316 base Sma I fragment of M-MulV DNA was labeled with ^{32}P. This probe allows for detection of the 3 kb internal fragment of Bam HI–digested M-MuLV together with other sequences derived from endogenous viruses (Figure 4, lanes a, d, g). Uninfected EC-A1 cells lacked the 3 kb M-MuLV-specific sequence (Figure 4, lane a). DNA of EC-A1 cells 48 hr after infection contained the 3 kb band (Figure 4, lane g), and this fragment was not derived from the unintegrated M-MuLV, as unintegrated proviral DNA disappeared from the Hirt supernatant fraction of EC-A1 cells before 48 hr of infection. Double digestion of 48 hr postinfection DNA with Hap II and with Msp I completely eliminated the 3 kb Bam HI band, suggesting that this part of the M-MuLV genome is devoid of DNA methylation (Figure 4, lanes h, i). DNA was isolated from EC-A1(Mo)4 cells that had undergone multiple cell cycling after virus infection. The M-MuLV-specific 3 kb Bam HI band was clearly demonstrable in the cellular DNA (Figure 4, lane d). However, the same sequence was now resistant to Hap II digestion (Figure 4, lane e). Other bands derived from endogenous virus genomes were also resistant to Hap II (Figure 4, lanes b, e, h). DNAs from another EC-A1(Mo) clone, EC-A1(Mo)22, and randomly infected EC-A1 cells passaged at least 20 times also contained the Hpa II–resistant 3 kb Bam HI fragment specific to M-MuLV.

Experiments were performed to determine the time of the DNA methylation of M-MuLV genome after infection. DNA was isolated from M-MuLV-infected EC-A1 cells on days 4, 10 and 15 after infection, and analyzed for the state of DNA methylation (Figure 5). The M-MuLV genome was devoid of methylation at

Hap II sites up to day 10 (Figure 5, lanes b, e). However, the viral genome was resistant to Hap II when it was isolated from 15-day-old cultures. Therefore, the integrated M-MuLV genome was methylated between days 10 and 15 in culture. Production of M-MuLV from infected EC-A1 cells during these 10 days was nil.

These results indicate that although unintegrated and newly integrated M-MuLV genomes were devoid of DNA methylation, the same sequence is methylated in the undifferentiated cells kept in culture for over 15 days.

Lack of Change in the State of DNA Methylation of the M-MuLV Genome during Differentiation of Teratocarcinoma Cells

DNA was isolated from undifferentiated EC-A1(Mo)4 cells and dimethylacetamide-induced differentiated EC-A1(Mo)4 cells. These DNAs contained the 3 kb Bam HI fragment of the integrated M-MuLV genome, as well as other bands derived from endogenous virus genomes (Figure 6). The 3 kb fragment of M-MuLV in undifferentiated EC-A1(Mo)4 cells was again resistant to digestion with Hap II, confirming the result in Figure 4, lane e (Figure 6, lane b). Similar resistance to Hap II enzyme was noted when the DNA from differentiated EC-A1(Mo)4 cells was analyzed. Therefore, the pattern of DNA methylation of at least the 3 kb Bam HI fragment did not change during differentiation of the cells, yet inducibility of the viral genome by 5-AzaCyd treatment changed drastically.

Transfection with DNA from M-MuLV Infected Teratocarcinoma Cells

DNA could be isolated from EC-A1 cells 2 days after infection with M-MuLV and also from EC-A1(Mo) clones. SC-1 cells were transfected with these DNAs (Table 3). Although at a low efficiency, DNA from EC-A1 cells infected with M-MuLV 2 days previously was capable of producing M-MuLV, while DNA from EC-A1(Mo) clones was not.

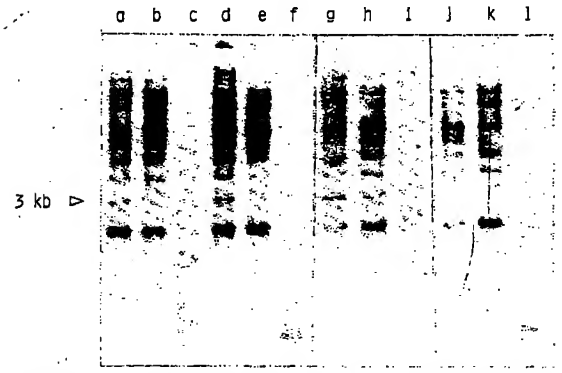

Figure 5. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from EC-A1 cells on the day 4 (lanes a, b and c), day 10 (lanes d, e and f) or day 15 (lanes g, h and i) after infection with M-MuLV, or from uninfected EC-A1 cells (lanes j, k and l). (Lanes a, d, g and j) digestion with Bam HI; (lanes b, e, h and k) digestion with Bam HI plus Hap II; (lanes c, f, i and l) digestion with Bam HI plus Msp I.

Discussion

Jaenisch and coworkers have demonstrated that endogenous virus can be formed by infection of the preimplantation stage embryo cells, with exogenous virus (Jaenisch et al., 1975; Jaenisch, 1976). Infection of undifferentiated stem cells of mouse teratocarcinoma leads to silencing of the exogenously acquired

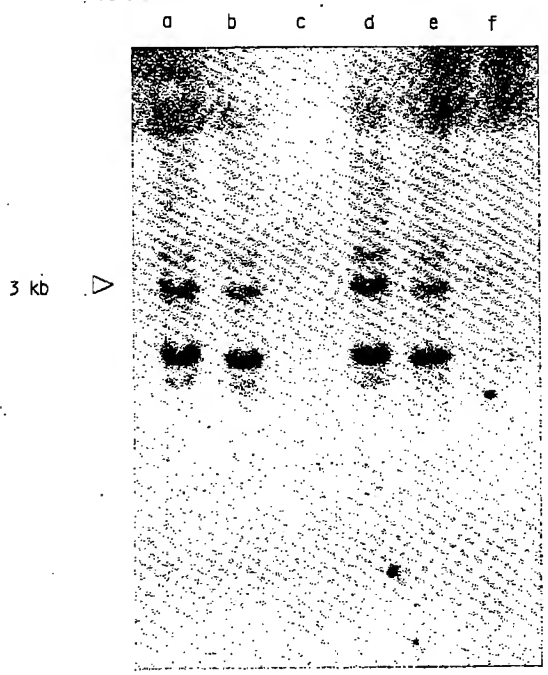

Figure 6. Blot Hybridization of M-MuLV Genome in Undifferentiated and Differentiated EC-A1(Mo)4 Cells

DNA was isolated from undifferentiated (lanes a, b and c) and differentiated EC-A1(Mo)4 cells (lanes d, e and f). The probe used was a ^{32}P-labeled 1.65 kb Bam HI–Hind III fragment of M-MuLV DNA. (Lanes a and d) digestion with Bam HI; (lanes b and e) digestion with Bam HI plus Hap II; (lanes c and f) digestion with Bam HI plus Msp I.

Table 3. Transfection of SC-1 Cells with DNA from M-MuLV-Infected Cells

DNA Source	DNA per Dish (μg)	Virus-Positive/ Total
SC-1	20	0/20
SC-1 M-MuLV-infected	20	23/30
EC-A1	20	0/20
EC-A1(Mo)4	20	0/20
EC-A1(Mo)22	20	0/20
EC-A1(Mo)26	20	0/20
EC-A1 M-MuLV-infected[a]	20	9/20

[a] DNA was isolated 2 days after infection with M-MuLV.

type C virus genome (Teich et al., 1977; Speers et al., 1980). However, the mechanism of repression of integrated viral genome in the cells of the preimplantation stage embryo and in teratocarcinoma cells has remained unknown.

Undifferentiated teratocarcinoma cells have characteristics of preimplantation stage embryo cells in that they lack H-2 antigen on the cell surface (Artzt and Jacob, 1974), possess two active X chromosomes in female cells (Martin et al., 1978; McBurney and Strutt, 1980) and have the potential to differentiate into a variety of cell types (Kleinsmith and Pierce, 1964; Mintz and Illmensee, 1975). Thus teratocarcinoma cells are a pertinent model for studies on the undifferentiated state of embryogenesis.

For cells to be in an undifferentiated state, they must be equipped with the potential to suppress specifically expression of genes required only for differentiated cells. Mouse teratocarcinoma cells are resistant to exogenously incorporated genetic elements such as SV40 and polyoma virus (Swartzendruber and Lehman, 1975; Swartzendruber et al., 1977; Segal and Khoury, 1979) as well as to murine leukemia virus. These viral genomes may be regarded as unnecessary luxury genes in undifferentiated teratocarcinoma cells.

We found that the unintegrated M-MuLV genome and at least part of the newly integrated genome, the 3 kb Bam HI fragment, in undifferentiated teratocarcinoma cells are devoid of DNA methylation, yet the expression of the genome is repressed. SV40 DNA was also shown to be undermethylated in teratocarcinoma cells (Friedrich and Lehman, 1981). Treatment with retinoic acid, which triggers differentiation of the cells, can activate virus expression in undifferentiated cells, provided that drug treatment follows shortly after infection with M-MuLV, when methylation of the M-MuLV genome has not yet occurred (Table 1). However, the frequency of virus expression induced by retinoic acid was rather low, and for the entire course of differentiation about 10 days were required. The low frequency thus observed might be due to the incompleteness of differentiation. The finding that the freshly integrated M-MuLV genome in undifferentiated teratocarcinoma cells is transfectionally active on SC-1 cells (Table 3) suggests that this genome could be transcriptionally active in differentiated cells. Therefore, the M-MuLV genome freshly acquired by EC-A1 cells seems to be devoid of DNA modification, and DNA modification is known to reduce the rate of transcription (Stuhlmann et al., 1981; Hoffmann et al., 1982). Actually, the M-MuLV genome was devoid of methylation at Hpa II sites for at least 10 days after infection; it was not expressed in the undifferentiated cells. The 3 kb fragment of M-MuLV genome is subsequently methylated in infected cells kept in culture for over 15 days. The virus genome becomes transfectionally inactive and cannot be induced by simple differentiation of the cells. Also, suppression of DNA methylation by 5-AzaCyd did not activate expression of the virus genome in EC-A1(Mo) clones. Dot blot hybridization analysis of RNA isolated from 5-AzaCyd-treated EC-A1(Mo)4 cells that carry the methylated M-MuLV genome did not contain RNA hybridizable to M-MuLV probe, while in the BrdUrd-treated culture there was a marked increase of M-MuLV transcript (O. Niwa, unpublished observation). Therefore, we conclude that the M-MuLV genome is repressed in the undifferentiated cells by a mechanism other than DNA methylation. This repression can be unblocked by the treatment of the cells with BrdUrd. EC-A1 cells freshly infected with M-MuLV, and carrying the unmethylated viral genome, and EC-A1(Mo) clones presumably carrying the methylated M-MuLV genome were both induced by BrdUrd treatment (see Table 1 and above). Incorporation of BrdUrd into DNA does not affect the level of methylation (unpublished observations).

DNA methylation has no effect on the transcriptional activity of the M-MuLV genome in undifferentiated cells. The M-MuLV genome is nevertheless methylated in cells that have undergone multiple cell cycling. Genes that are not transcribed may be preferentially methylated in undifferentiated cells.

The pattern of DNA methylation of the 3 kb Bam HI fragment did not change during differentiation of EC-A1(Mo)4 cells. However, the M-MuLV genome is now susceptible to induction by 5-AzaCyd. Repression of the viral genome by DNA methylation thus seems operative only in differentiated cells. DNA methylation is thought to suppress gene expression through a condensation of chromatin, and antisera raised against methylcytosine bind to the heterochromatic region of the mouse chromosome (Miller et al., 1974). Methylcytosine is more abundant in the fraction of chromation that is resistant to DNAase digestion (Razin and Ceder, 1977). Methylated sequences of endogenous viruses are located on chromatin regions that are resistant to DNAase I (van der Putten et al., 1982).

Cells of the preimplantation stage embryo have two active X chromosomes and lack heterochromatin (Epstein et al., 1978). Heterochromatinization of an X chromosome occurs during differentiation of teratocarcinoma cells (Martin et al., 1978; McBurney and Strutt, 1980). Condensation of the heterochromatic region may even be facilitated by the presence of yet unidentified chromatin protein(s), and differentiation of cells may trigger production of this chromatin protein, which inactivates already methylated luxury genes by condensation of their chromatin regions.

In light of all these data, we propose the presence of two independent mechanisms regulating gene expression in mammalian cells. The first mechanism which is operating in undifferentiated cells, is not influenced by the state of DNA methylation for its function. Repression of transcription by this mecha-

nism may involve discrimination against the luxury gene promoter. Host range mutants of polyoma virus that can replicate on undifferentiated teratocarcinoma cells were found to possess mutations at the promoter region (Sekikawa and Levine, 1981). The second mechanism suppresses the expression of methylated genes by changing the conformation of their chromatin domains, and this mechanism operates only in differentiated cells. When M-MuLV infects undifferentiated teratocarcinoma cells, the viral genome is suppressed by the first mechanism. The transcriptionally inactive genome of M-MuLV may be gradually methylated during replication of the host cells. Differentiation of the cells terminates regulation by the first mechanism and activates the second mechanism, which now recognizes the methylated genome of M-MuLV and represses its expression by condensation of the chromatin domain. 5-AzaCyd unblocks the second mechanism by decreasing the level of methylcytosine, while BrdUrd unblocks both mechanisms. Since BrdUrd is known to change the affinity of DNA-binding proteins for DNA (Lin and Riggs, 1972, 1976; Goeddel et al., 1977), binding of some chromatin protein(s) to the promoter region of M-MuLV may be responsible for operation of the first mechanism. The first mechanism of the repression of gene expression seems to be *trans*-acting, since the extrachromosomal genome of SV40 and the M-MuLV genome integrated at random sites in undifferentiated cells are repressed. These findings suggest that the first mechanism may inactivate gene expression, possibly by some diffusible repressor-like protein, and if such is the case, the *trans*-acting nature can be readily explained.

During preparation of this manuscript, a report appeared on a similar subject (Stewart et al., 1982) These authors' data suggest that the M-MuLV genome becomes methylated immediately after integration. This difference between our results and theirs might relate to different cell lines. We used PCC4-derived EC-A1 cells, while they used F-9 cells. Gautsch and Wilson (1983) apparently obtained findings similar to those we report here.

Experimental Procedures

Cell Lines and Viruses

EC-A cells (Gautsch, 1980), a subline of PCC4 Aza1 cells (Jakob et al., 1973), were kindly provided by J. Gautsch. EC-A1 cells, a subclone of EC-A cells isolated in our laboratory, were also used. F-9 cells (Strickland and Mahdavi, 1978) were obtained from K. Sekigawa. SC-1 cells, a mouse embryo fibroblast line derived from a feral mouse (Hartley and Rowe, 1975), were obtained from A. Declève. Teratocarcinoma cell lines were grown in Eagle's minimal essential medium supplemented with 10% heat-inactivated fetal calf serum. SC-1 cells were grown in MEM alpha medium (Gibco) supplemented with 10% heat-inactivated calf serum.

M-MuLV was obtained from A. Ishimoto and grown on SC-1 cells.

Virus Infection, Transfection and Biological Assays

Procedures of virus infection were as described previously (Niwa et al., 1973). Titration of M-MuLV stock was on SC-1 cells by UV-XC assay (Rowe et al., 1970). M-MuLV-infected teratocarcinoma cells were assayed directly or after cocultivation with SC-1 cells by the reverse XC cell assay (Niwa et al., 1973).

Transfection of SC-1 cells with DNA from M-MuLV-infected EC-A1 cells was carried out by a procedure described by other workers (Stuhlmann et al., 1981).

Drug Treatments

Stock solutions of retinoid and BrdUrd were prepared in dimethylsulfoxide at concentrations of 10 mM and 1 mg/ml, respectively, and kept in the dark at −20°C. Teratocarcinoma cells were incubated with these drugs for 24 hr at 37°C either before or after infection with M-MuLV. When cocultivation with SC-1 cells was used for the amplification of progeny virus, teratocarcinoma cells were treated for 30 min with 25 μg/ml mitomycin C to suppress overgrowth of the undifferentiated cells.

Differentiated teratocarcinoma cells were obtained by the procedure of Speers et al. (1980) with slight modification. Undifferentiated teratocarcinoma cells grown as monolayer cultures were nourished every 3 days with a medium containing 10 mM dimethylacetamide for 2 weeks. Cultures consisted only of cells with an epithelial morphology. Although retinoic acid also induced differentiation of EC-A1 cells, it was more cytotoxic than dimethylacetamide.

Base Analysis

The level of methylcytosine in the newly synthesized DNA was assayed as described previously (Niwa and Sugahara, 1981).

Extraction of Cellular RNA and DNA

Total cellular RNA was isolated by sedimentation through cesium chloride as described by others (Chirgwin et al., 1979).

For the isolation of DNA, dishes were washed with phosphate-buffered saline solution and digested at 37°C for 2 hr with 100 μg/ml RNAase A in 1% sodium dodecylsulfate, 0.1 M NaCl, 5 mM EDTA and 20 mM Tris–HCl (pH 8.0). Proteinase K was then added at 100 μg/ml, and dishes were incubated for another 2 hr at 37°C. The lysate was extracted three times with phenol–chloroform and ethanol-precipitated. DNA thus extracted was used for further analysis. Unintegrated viral DNA was isolated by the procedure of Hirt (1967).

Restriction Endonuclease Digestion, Gel Electrophoresis and Hybridization

Restriction enzymes were obtained from Takara Shuzo Co., Ltd. (Kyoto, Japan) except for Msp I, which was purchased from New England Biolaboratory. After cleavage with restriction enzymes, DNA samples were electrophoresed on 0.7% horizontal agarose gels and transferred to a nitrocellulose sheet, as described by Southern (1975).

Total cellular RNA was denatured, as described (McMaster and Carmichael, 1977), electrophoresed on 1.1% agarose gels and transferred to DBM paper accordingly (Alwine et al., 1977). For the dot blot hybridization, nondenatured RNA was spotted directly onto nitrocellulose filter (Thomas, 1980).

The recombinant plasmid which carries the 8.2 kb M-MuLV genome at the Hind III site and which was cloned by J. W. Hoffmann, was a generous gift from R. A. Weinberg. The DNA fragment of 8.2 kb containing the entire sequence of M-MuLV was purified by agarose gel electrophoresis. M-MuLV DNA was further digested by Bam HI and Sma I. The Hind III–Bam HI fragment of 1.65 kb was recovered from agarose gel after electrophoresis. The Sma I fragment of 316 bases was electrophoresed through a polyacrylamide gel. Recombinant plasmids carrying 18S and 28S rRNA gene of the mouse were a kind gift from R. Kominami (Urano et al., 1980; Kominami et al., 1982). These DNAs were labeled by the nick translation procedure (Maniatis et al., 1975). Specific activities of the probes were 1–4 × 10^8 cpm/μg.

Acknowledgments

We thank Drs. R. A. Weinberg and J. W. Hoffmann for the generous gift of recombinant plasmid, M. Yoshida for pertinent advice on biochemical techniques, A. Ishimoto for providing M-MuLV stock,

Prof. S. Numa, in whose laboratory part of this investigation was carried out; R. Kominami for provision of plasmids carrying rRNA genes, K. Sekikawa from whom we obtained the F-9 cells, M. Ohara for reading the manuscript and Y. Fukuda for secretarial services. This work was supported by a grant-in-aid from the Ministry of Education, Science, and Culture, Japan.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 14, 1982; revised January 25, 1983

References

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA *74*, 5350–5354.

Artzt, K. and Jacob, F. (1974). Absence of serologically detectable H-2 on primitive teratocarcinoma cells in culture. Transplantation *17*, 632–634.

Bernstine, E. G., Hooper, M. L., Grancharnp, S. and Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. Proc. Nat. Acad. Sci. USA *70*, 3899–3903.

Chan, H. W., Bryan, T., Moore, J. L., Staal, S. P., Rowe, W. P. and Martin, M. A. (1980). Identification of ecotropic proviral sequences in inbred mouse strains with a cloned subgenomic DNA fragment. Proc. Nat. Acad. Sci. USA *77*, 5779–5783.

Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. and Rowe, W. P. (1974). Qualitative and quantitative studies on AKR-type murine leukemia virus sequences in mouse DNA. Cold Spring Harbor Symp. Quant. Biol. *39*, 1085–1101.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Dolberg, D. S., Bacheler, L. T. and Fan, H. (1981). Endogenous type C retroviral sequences of mice are organized in a small number of virus-like classes and have been acquired recently. J. Virol. *40*, 96–106.

Epstein, C. J., Smith, S., Travis, B. and Tucker, G. (1978). Both X chromosomes function before visible X-chromosome inactivation in female mouse embryos. Nature *274*, 500–503.

Friedrich, T. D. and Lehman, J. M. (1981). The state of simian virus 40 DNA in the embryonal carcinoma cells of the murine teratocarcinoma. Virology *110*, 159–166.

Gautsch, J. W. (1980). Embryonal carcinoma stem cells lack a function required for virus replication. Nature *285*, 110–112.

Gautsch, J. W. and Wilson, M. (1983). Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature, in press.

Goeddel, D. V., Yansura, D. G. and Caruthers, M. H. (1977). Binding of synthetic lactose operator DNAs to lactose repressors. Proc. Nat. Acad. Sci. USA *74*, 3292–3296.

Groudine, M., Eisenman, R. and Weintraub, H. (1981). Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature *292*, 311–317.

Hartley, J. W. and Rowe, W. P. (1975). Clonal cell line from a feral mouse embryo which lacks host range restriction for murine leukemia virus. Virology *65*, 128–134.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365–369.

Hoffman, J. W., Steffen, D., Gusella, J., Tabin, C., Bird, S., Cowing, D. and Weinberg, R. A. (1982). DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. *44*, 144–157.

Heubner, K., Tsuchida, N., Green, C. and Croce, C. M. (1979). A murine teratocarcinoma stem cell line carries suppressed oncogenic virus genomes. J. Exp. Med. *150*, 392–405.

Ihle, J. N., Joseph, D. R. and Domotor, J. J. (1979). Genetic linkage of C3H/HeJ and BALB/c endogenous ecotropic C-type viruses to phosphoglucomutase-1 on chromosome 5. Science *204*, 71–73.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc. Nat. Acad. Sci. USA *73*, 1260–1264.

Jaenisch, R., Fan, H. and Croker, B. (1975). Infection of preimplantation mouse embryos and of new born mice with leukemia virus: tissue distribution of viral DNA and RNA leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA *72*, 4008–4012.

Jakob, H., Boon, T., Gillard, J., Nicholas, J. F. and Jacob, F. (1973). Teratocarcinoma de la souris: isolement, culture et proprietes de cellules a potentialites multiples. Ann. Microbiol. (Paris). *124B*, 269–282.

Kleinsmith, L. J. and Pierce, G. B., Jr. (1964). Multipotentiality of single embryonal carcinoma cells. Cancer Res. *24*, 1544–1551.

Kominami, R., Mishima, Y., Urano, Y., Sakai, M. and Muramatsu, S. (1982). Cloning and determination of the transcription termination site of ribosomal RNA gene of the mouse. Nucl. Acids Res. *10*, 1963–1979.

Kozak, C. A. and Rowe, W. P. (1979). Genetic mapping of the ecotropic murine leukemia virus-inducing locus of BALB/c mouse to chromosome 5. Science *204*, 69–71.

Kozak, C. A. and Rowe, W. P. (1980). Genetic mapping of the ecotropic virus-inducing locus *Akv*-2 of the AKR mouse. J. Exp. Med. *152*, 1419–1423.

Lehman, J. M., Speers, W. C., Swartzendruber, D. E. and Pierce, G. B. (1974). Neoplastic differentiation: characteristics of cell lines derived from a murine teratocarcinoma. J. Cell. Physiol. *84*, 13–28.

Lin, S. Y. and Riggs, A. (1972). *Lac* operator analogues: bromodeoxyuridine substitution in the *lac* operator affects the rate of dissociation of the *lac* repressor. Proc. Nat. Acad. Sci. USA *69*, 2574–2576.

Lin, S. Y. and Riggs, A. (1976). The binding of *lac* repressor and the catabolite gene activator protein to halogen-substituted analogues of poly (d(A-T)). Biochim. Biophys. Acta. *432*, 185–191.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Martin, G. R., Epstein, C. J., Travis, B., Tucker, G., Yatziv, S., Martin, D. M., Clift, S. and Cohen, S. (1978) X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro. Nature *271*, 329–333.

McBurney, M. W. and Strutt, B. J. (1980). Genetic activity of X chromosomes in pluripotent female teratocarcinoma cells and their differentiated progeny. Cell *21*, 357–364.

McMaster, G. K. and Carmichael, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agrose gels by using glyoxal and acridine orange. Proc. Nat. Acad. Sci. USA *74*, 4835–4838.

Miller, O. J., Schnedl, W., Allen, J. and Erlanger, B. F. (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature *251*, 636–637.

Mintz, B. and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc. Nat. Acad. Sci. USA *72*, 3585–3589.

Niwa, O. and Sugahara, T. (1981). 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. Proc. Nat. Acad. Sci. USA *78*, 6290–6294.

Niwa, O., Declerve, A., Lieberman, M. and Kaplan, H. S. (1973). Adaptation of plaque assay methods to the in vitro quantitation of the radiation leukemia virus. J. Virol. *12*, 68–73.

Peries, J., Alves-Cardoso, E., Canivet, M., Devons-Guillemin, M. C. and Lasneret, J. (1977). Lack of multiplication of ecotropic murine C-type viruses in mouse teratocarcinoma primitive cells. J. Nat. Cancer Inst. *59*, 463–465.

Razin, A. and Ceder, H. (1977). Distribution of 5-methylcytosine in chromatin. Proc. Nat. Acad. Sci. USA *74*, 2725–2728.

Rowe, W. P., Pugh, W. E. and Hartley, J. W. (1970). Plaque assay techniques for murine leukemia viruses. Virology *42*, 1136–1139.

Rowe, W. P., Hartley, J. W. and Besmer, T. (1972). Genetic mapping of a murine leukemia virus–inducing locus of AKR mice. Science *178*, 860–862.

Segal, S. and Khoury, G. (1979). Differentiation as a requirement for simian virus 40 gene expression in F-9 embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *76*, 5611–5615.

Sekikawa, K. and Levine, J. (1981). Isolation and characterization of polyoma host range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *78*, 1110–1104.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–518.

Speers, W. C., Gautsch, J. W. and Dixon, F. J. (1980). Silent infection of murine embryonal carcinoma cells by Moloney murine leukemia virus. Virology *105*, 241–244.

Stewart, C. L., Stuhlman, H., Jähner, D. and Jaenisch, R. (1982) *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *79*, 4098–4102.

Strickland, S. and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell *15*, 393–403.

Stuhlmann, H., Jähner, D. and Jaenisch, R. (1981). Infectivity and methylation of retrovirus genomes is correlated with expression in the animal. Cell *26*, 221–232.

Swartzendruber, D. E. and Lehman, J. M. (1975). Neoplastic differentiation: interaction of simian virus 40 and polyoma virus with murine teratocarcinoma cells in vitro. J. Cell Physiol. *85*, 179–188.

Swartzendruber, D. E., Friedrich, T. D. and Lehman, J. M. (1977). Resistance of teratocarcinoma stem cells to infection with simian virus 40: early events. J. Cell Physiol. *93*, 25–30.

Teich, N. M., Weiss, R. A., Martin G. R. and Lowy, D. R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973–982.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. USA *77*, 5201–5205.

Urano, Y., Kominami, R. and Muramatsu, S. (1980). The nucleotide sequence of the putative transcription initiation site of a cloned ribosomal RNA gene of the mouse. Nucl. Acids Res. *8*, 6043–6058.

van der Putten, H., Quint, W., Verma, I. M. and Berns, A. (1982). Moloney murine leukemia virus–induced tumors: recombinant proviruses in active chromatin regions. Nucl. Acids Res. *10*, 577–592.

TAB 6

ELSEVIER

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Gene 261 (2000) 181–188

www.elsevier.com/locate/gene

Retroviruses in foreign species and the problem of provirus silencing☆

J. Svoboda[a,*], J. Hejnar[a], J. Geryk[a], D. Elleder[a], Z. Vernerová[b]

[a]*Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, 166 37 Prague 6, Czech Republic*
[b]*Department of Pathology, First Faculty of Medicine, Charles University, Prague, Czech Republic*

Received 29 May 2000; received in revised form 18 August 2000; accepted 5 October 2000
Received by G. Bernardi

Abstract

Retroviruses are known to integrate in the host cell genome as proviruses, and therefore they are prone to cell-mediated control at the transcriptional and posttranscriptional levels. This plays an important role especially after retrovirus heterotransmission to foreign species, but also to differentiated cells. In addition to host cell-mediated blocks in provirus expression, also so far undefined host specificities, deciding upon the pathogenic manifestation of retrovirus heterotransmission, are in play. In this respect, we discuss especially the occurrence of wasting disease and immunodeficiency syndrome, which we established also in avian species using avian leukosis virus subgroup C (ALV-C) inoculated in mid-embryogenesis in duck or chicken embryos. The problem of provirus downregulation in foreign species or in differentiated cells has been in the recent years approached experimentally. From a series of observations it became apparent that provirus downregulation is mediated by its methylation, especially in the region of proviral enhancer-promoter located in long terminal repeats (LTR). Several strategies have been devised in order to protect the provirus from methylation using LTR modification and/or introducing in the LTR sequence motifs acting as antimethylation tags. In such a way the expression of retroviruses and vectors in foreign species, as well as in differentiated cells, has been significantly improved. The complexity of the mechanisms involved in provirus downregulation and further possibilities to modulate it are discussed. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Retrovirus heterotransmission; Retroviral vector; Methylation; CpG; Long terminal repeat(s); Histone deacetylase

1. Introduction

The topic of this review is focused on selected aspects of retrovirus heterotransmission, but also touches the problem of retroviral genome silencing in differentiated and other types of cells. In case of heterotransmission, many host cell factors are in play, which decide whether or to what degree the retroviral genome will be expressed and what pathogenic consequences may be triggered. As is known generally, the retroviral genome becomes integrated in the cell genome as a provirus and, therefore, it is not surprising that it is highly influenced by the host-cell gene-regulation machinery. In discussing such downregulation of the provirus we are in fact dealing with post-integration blocks in provirus expression, the nature of which is epigenetic and mediated by new host cell factors. The unusual cell milieu, in concert with which the virus has not been evolving, can be lacking some factors like those enabling viral RNA export from the nucleus, or provide unusual factors like those changing the viral RNA splicing. However, of main importance is cell transcriptional regulation, which in many cases leads to provirus silencing. It is, therefore, not surprising that in phylogenetically distant host cells the provirus can integrate, but in many cases does not produce an infectious progeny. We call such host cells non-permissive, in contrast to permissive cells where formation of infectious virions takes place.

Non-permissivenes to retroviral infection has been for a

Abbreviations: ALV, avian leukosis virus(es); ALV-B, ALV-C, ALV-D, avian leukosis virus subgroup B, C, and D, respectively; *aprt*, adenosine-phosphoribosyltransferase gene; β-*geo*, fused β-galactosidase and neomycin resistance gene; BLV, bovine leukemia virus; CAT, chloramphenicol acetyltransferase; cHS4, chicken hypersensitive site 4; EC, embryonic carcinoma; GFP, green fluorescence protein; HDAC, histone deacetylase; HIV, human immunodeficiency virus; IFN-SAR, interferon scaffold attachment region; LCR, locus control region(s); LTR, long terminal repeat(s); MeCP, methyl-CpG-binding protein; MEL, murine erythroleukemia; MLV, murine leukemia virus(es); MSV, murine sarcoma virus; NCR, negative control region; *neo*, neomycin resistance gene; PR RSV, Prague strain of RSV; RSV, Rous sarcoma virus; SIV, simian immunodeficiency virus; SIVcpz, chimpanzee SIV; SIVsm, sooty mangabey SIV; TSA, trichostatin; X-MLV, xenotropic MLV

☆ Extended version of the lecture "Modification of retrovirus pathogenicity by transspecies transfer" presented at the workshop "Microbial Variation and Evolution" at Ischia, Italy (organizers G. Bernardi and B. Fantini).

* Corresponding author. Tel.: +42-2-243-10-238; fax: +42-2-243-10-955.

E-mail address: svoboda@img.cas.cz (J. Svoboda).

PII: S0378-1119(00)00481-9

long time ascribed to foreign species host cells. Interestingly enough, non-permissiveness to retroviral infection marks also some cells of the virus species origin. This has been recognized especially in cases of cultured differentiated cells. In such a way, surmounting hurdles negatively influencing provirus expression became of importance for an efficient application of retroviral vectors for gene therapy of developmentally committed stem cells and differentiated cells. Therefore, we also discuss the provirus fate in some defined cases of differentiated cells that originate from the same species as the retrovirus. There are common features shared among all these situations, pointing to a role of cell-mediated gene silencing as an important factor that superimposes upon provirus expression.

2. Trans-class retrovirus heterotransmission

This extreme situation was first achieved in the case of chicken Rous sarcoma virus (RSV) strains transmitted first to rodents and later also to other mammalian species, including monkeys (rev. Svoboda, 1986). RSV genome expression in mammalian cells is governed by a series of factors, especially by flanking DNA sequences, their richness in GC (Fincham and Wyke, 1991; Rynditch et al., 1991), but also posttranscriptional steps are involved, which are blocked in mammalian cells (rev. Svoboda, 1998). More recent progress indicates that of importance are RSV LTR, which are prone to methylation in mammalian cells.

It was shown experimentally that a reporter gene driven by in vitro methylated RSV LTR is more efficiently suppressed in mammalian cells as compared to chicken cells (Hejnar et al., 1999). In spite of the fact that there are 16 CpGs in the Prague strain of Rous sarcoma virus (PR RSV) LTR and multiple CpGs are present in leader sequences, methylation of one CCGG *Hpa*II site, located downstream of the promoter region but close to the single provirus transcriptional start, was sufficient to produce dramatic reporter downregulation in mammalian cell lines. It should be noted that these experiments were done using transient transfection of methylated proviral DNAs and unmethylated controls. Expression of unmethylated LTR was comparable in both avian and non-permissive mammalian cells, suggesting that both types of cells harbor sufficient transcriptional machinery required by RSV LTR. Silencing of RSV proviruses is therefore a post-integration event.

The significance of LTR methylation for provirus downregulation has been recently approached using two strategies. The first one, schematically shown in Fig. 1, implies insertion of four canonical Sp1 binding sites in the RSV LTR enhancer region using *Eco*RI sites. Such reconstruction has been documented and discussed in detail (Machoň et al., 1998), and it was found that Sp1 insertion significantly increases LTR-driven chloramphenicol acetyltransferase (CAT) reporter gene expression, especially in hamster cells using both transient and stable transfection assays. This is in agreement with findings revealing that Sp1 binding sites represent the critical part of sequences acting as antimethylation tags, as was documented clearly in the case of a CpG island containing 1.7 kb DNA located in front of the adenosine-phosphoribosyltransferase (*aprt*) gene (Mummaneni et al., 1993; Brandeis et al., 1994; Macleod et al., 1994). Acquisition of an Sp1 binding site by mutation in the LTR enhancer region was shown previously to activate MLV (murine leukemia virus) transcription in embryonic carcinoma (EC) cells (Prince and

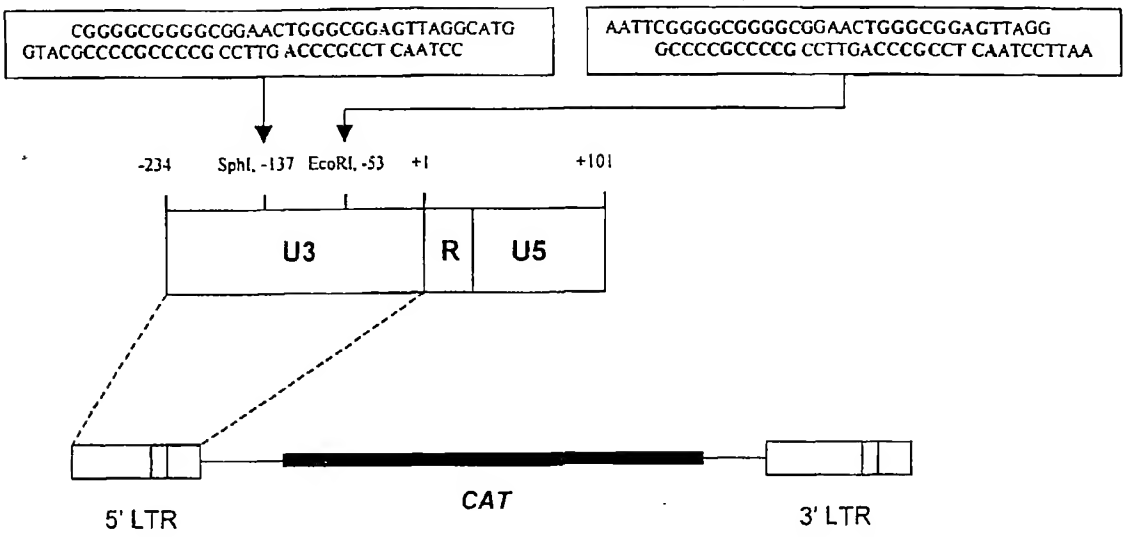

Fig. 1. Scheme of the CAT reporter vector employed for insertion of Sp1 sites into LTR sequence and transfection experiments. Sites of insertion (*Sph*I and *Eco*RI) are denoted by arrows. Sp1 binding sites within the inserted sequences are in bold.

Rigby, 1991). In a similar way, transcriptional activation of human endogenous retrovirus (HERV-H) was accomplished by Sp1 occurrence in either LTR enhancer (Nelson et al., 1996) or promoter (Sjøttem et al., 1996; Anderssen et al., 1997). The second approach is based on the incorporation of the mouse *aprt* gene CpG island immediately upstream to the RSV LTR-driven fused β-galactosidase and neomycin resistance (β-*geo*) proviral reporter (see Fig. 2). After transfection of hamster cells and selection for neomycin resistance, such construction ensured stable transcriptional activity in a reasonable number of cell clones, whereas unprotected proviral reporters are inactivated. This effect can be attributed to the antimethylation protection of the CpG island for two reasons; first, the active proviruses remained unmethylated within the 5′ LTR as evidenced by the bisulphite sequencing technique, second, the transcription efficiency of the RSV LTR is not increased in the presence of the CpG island.

Thus, both approaches led to the conclusion that proviral LTR could be protected from DNA methylation in the foreign species host. Optimization of these protective strategies might open the way to construct improved RSV-based vectors for gene transfer, more suitable for expression in mammalian cells and without the risk of infectious retroviral progeny. Retroviral vector producer cells (helper cells) are an important object for such protection as well. These cells have been shown genetically unstable due to the methylation of integrated helper proviral constructions. Designing a helper virus to overcome cellular DNA methylation may therefore improve vector production (Young et al., 2000).

There exists also another way of trans-species retrovirus transmission utilizing xenotropic murine leukemia viruses (X-MLV) (rev. Levy, 1978). It is interesting that these viruses can replicate in some avian cells such as duck cells, but not in others like chicken cells (Levy, 1977). X-MLV can provide envelope components to pseudotype RSV virions and such pseudotypes transform and replicate in duck but not in chicken cells. In vivo inoculation of X-MLV or murine sarcoma virus (MSV) pseudotyped by X-MLV was performed in duck embryos or newborn ducklings (Levy et al., 1982). Evidence of virus persistence has been obtained, but not convincing data concerning their pathogenesis. In mammalian cells, X-MLV-pseudotyped RSV can also replicate in the presence of X-MLV first to a low titer, which increases with passages. Furthermore, an envelope component of ALV-C phenotypically mixed with X-MLV was detected after passaging in mammalian cells.

It is not known how X-MLV contributes to RSV replication in mammalian cells. Obviously, it provides at least a part of the Env glycoprotein required for penetration to mammalian cells. However, additional factors complementing the non-permissiveness of mammalian cells to RSV should be in play. Due to its ability to multiply in such cells, X-MLV can increase and facilitate some posttranscriptional steps such as viral RNA proper splicing and export from the nucleus and/or further steps involving cleavage of viral protein precursors and virion assembly. These facilitating effects of X-MLV should be synchronized with RSV infection, because superinfection of an already RSV-transformed mammalian cell, containing functional proviruses rescuable by fusion with chicken fibroblasts, does not result in virus production (Levy, 1977).

Because X-MLV contribution to RSV replication in mammalian cells increases with passages, a possibility of genetic exchange between these viruses should be taken into account. In every case, these questions should be reinvestigated using presently available efficient tools of molecular biology.

3. Retrovirus heterotransmission among species within the same class

One of the first thoroughly analyzed successful heterotransmissions was achieved by Duran-Reynals (1942), who showed that RSV produces both early and late appearing tumors when inoculated in young ducks. Since then, many

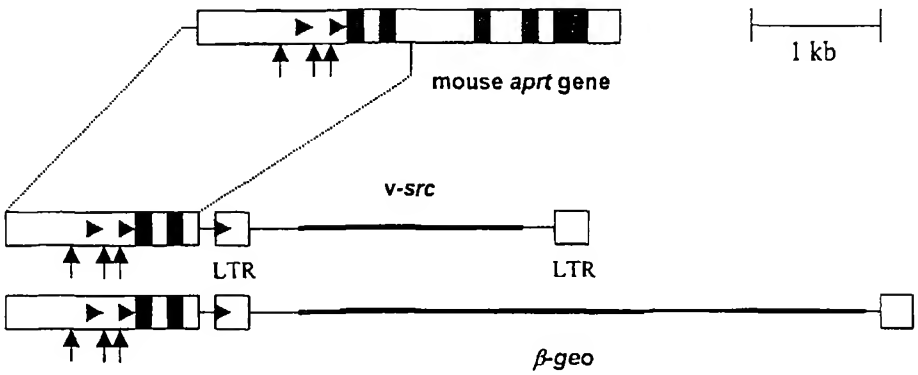

Fig. 2. Cloning of the *aprt* gene CpG island and RSV-based reporter proviruses. Filled boxes represent exons of the *aprt* gene. Vertical arrows denote the position of three Sp1 sites in the CpG island. Filled arrowheads denote transcription starts of the *aprt* gene and 5′ LTR.

Table 1
Trans-species retrovirus transmission

Virus	Species of origin	Original pathogenicity	Transmitted to species	New species pathogenicity	References
SIV_{cpz}	Chimpanzees	Zero to very low	Human	AIDS (HIV-1)	Gao et al., 1999; Weiss and Wrangham, 1999.
SIV_{sm}	Sooty mangabey	Zero to very low	Human	AIDS (HIV-2)	Hirsch et al., 1989; Gao et al., 1992; Sharp et al., 1995; Chen et al., 1996.
BLV	Cattle	Lymphocytosis, leukosis, B-tropic	Rabbit	Wasting disease, immunodeficiency	Burny et al., 1985; Altanerova et al., 1989; Wyatt et al., 1989; Kucerova et al., 1999.
Friend MLV	Mice	Erythroleukemia	Rat	Bone marrow suppression, $Thy1^+$ cell reduction	Mazgareanu et al., 1998.
ALV-C	Chicken	Anemia, probably immunodeficiency	Duck	Wasting disease, immunodeficiency, anemia	Karakoz et al., 1980;
ALV-B ALV-D		Anemia			Smith and Schmidt, 1982.

other retroviruses, including these of mammalian origin, were experimentally transmitted among different mammalian species. These transmissions were monitored mainly by virus oncogenic activity and virus persistence. Because of generally low retrovirus replication in foreign species, additional pathogenic virus activity, such as immunosuppression, usually do not appear. In many cases heterotransmission produced the same symptoms as in the species of the virus origin, but there are well-documented cases of changed virus pathogenicity. Further, we shall deal mainly with the symptoms of wasting disease accompanied by immunodeficiency resulting in increased susceptibility to various infectious agents.

The simian immunodeficiency viruses (SIV) highlight this situation. Generally, in the monkey species, in which SIV is indigenous, it replicates efficiently but does not produce any pathogenic changes. However, when transmitted to some other monkey species, SIV produces the immunodeficiency syndrome. Of special interest is SIV heterotransmission to humans. As is summarized in Table 1, good evidence based on molecular biology and epidemiology has been provided documenting that both human immunodeficiency virus 1 (HIV-1) and HIV-2 represent a consequence of respective SIV_{sm} (sooty mangabey SIV) and SIV_{cpz} (chimpanzee SIV) heterotransmission to humans. There is no doubt about the significance of this finding, which should be taken as a warning against potential danger of retrovirus transgression of species barriers associated with fulminating pathogenic changes. In order to understand these events, comparative data obtained with other members of the retrovirus family should be evaluated.

Interesting observations were made in the case of bovine leukemia virus (BLV) (Table 1). This virus responsible for cattle leukosis when transferred to newborn rabbits triggers clear symptoms of immunodeficiency. Because such a response was not found in other infected species, these observations indicate that rabbits respond to BLV inherently, in an unusual way.

In murine leukemia viruses, variants capable to produce immunodeficiency preferably have been isolated. This includes both Moloney MLV (Saha et al., 1994) and Friend leukemia complex (Faxvaag et al., 1993). In addition, as given in Table 1, Friend MLV transmission to newborn rats resulted in altered pathogenicity characterized by suppression of bone marrow cells, manifesting itself as reduced numbers of $Thy1^+$ cells.

Avian leukosis viruses have not been thoroughly investigated from the point of view of their immunopathogenicity in foreign avian species. As given in Table 1, ALV subgroup C were studied using intraembryonic inoculation both in chicken and ducks. According to data obtained so far, this subgroup produces symptoms of anemia in both species. However, in heterologous duck hosts, a fatal wasting disease together with conspicuous atrophy of the thymus tissue starting the first week after hatching was observed both by a decrease in the relative thymus to body weight, histologically characterized by clear depletion of the thymus cortical layer (Fig. 3) (Stepanets et al., 2000). Microscopically, bursa Fabricii, which constitutes a special B-cell producing organ in birds, also displayed the cortical layer depletion. In agreement with this observation, production of humoral antibodies against *Brucella abortus* antigens was significantly decreased in infected animals, which confirms that immunodeficiency is involved. The nature of this immunodeficiency is being investigated and the character of specific T- and other cell alteration in thymus is monitored by specific antibodies. In addition, changes in lymphoid organs in young chickens intraembryonally inoculated with different ALV subgroups should be investigated. According to our preliminary data, ALV-C does produce symptoms of thymus involution even in chickens. Thus ALV-C provides a suitable comparative system for establishing ways leading to the immunodeficiency syndrome both in homologous and heterologous hosts.

There is not a simple answer to the question why retroviruses in some species behave pathogenically or produce

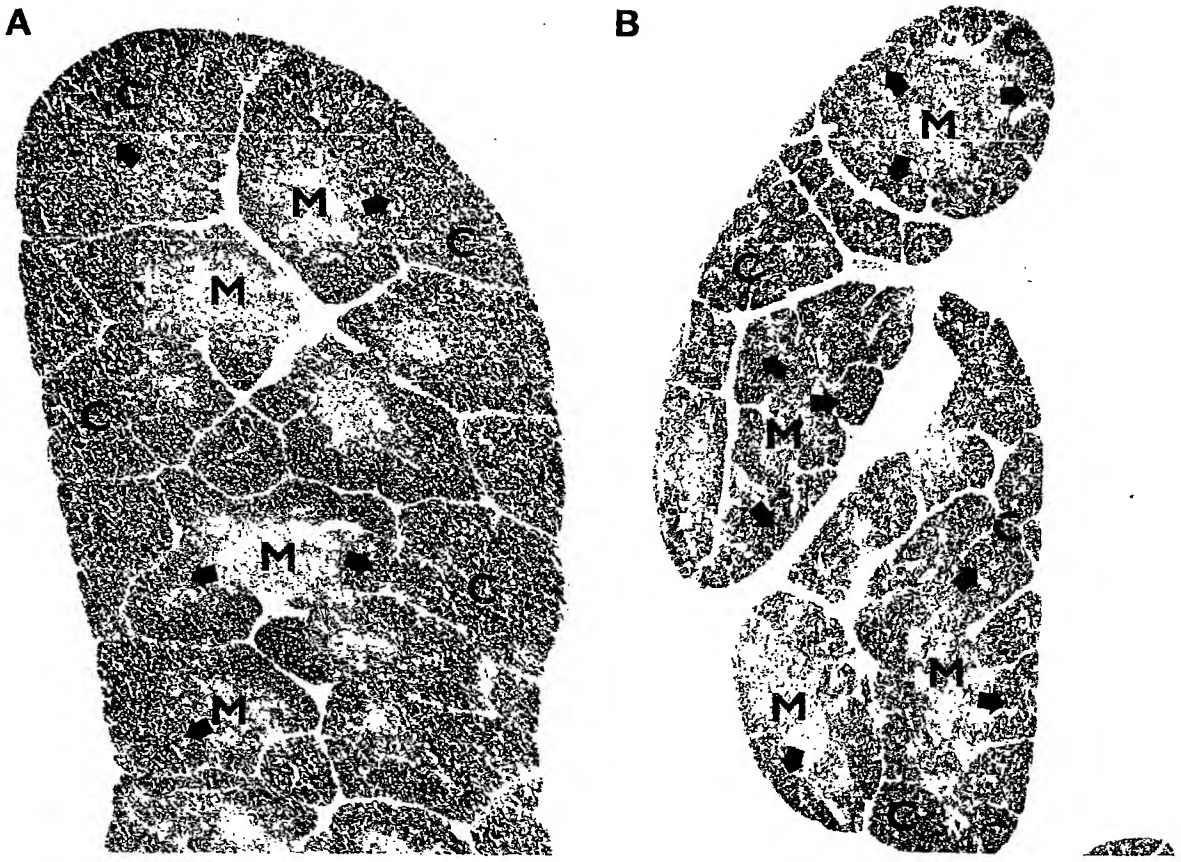

Fig. 3. Thymus sections of 7-day-old ducklings. (A) Thymus of a control animal, mock-injected with tissue culture medium. (B) Thymus of an animal infected in mid-embryogenesis with 10^5 ALV-C infectious virus. Thymic cortical layer is visibly depleted. C, cortex; M, medulla; arrowheads mark the interim between cortex and medulla. Stained with hematoxylin-eosin. Magnification 41 ×.

new or accentuated pathogenic symptoms. In some cases, such as Friend MLV, selection of virus mutants might be involved, but this seems not to be a general situation, because altered pathogenicity appears very soon after virus inoculation and is not correlated with certain specific virus gene alteration. Therefore other, especially host-specific factors should be taken into account. It is not known so far which of them play a decisive role. Retrovirus toxicity is in the first step related to the outcome of the viral envelope (product of the *env* gene) interaction with target cells, immune and bone-marrow cells included. Therefore, if an infected cell harbors unusually expressed receptors and co-receptors as well as factors required for further steps of virus penetration, they may undergo deterioration or apoptosis. As was discussed in relation to HIV (Fauci, 1996), the degree of cytokine activity and the outcome of interrelation among different cytokine pathways contribute to virus expression and its pathogenicity.

In spite of the fact that heterotransmission of SIV took place probably several times, we are lacking the exact information about this process. In contrast to the experimental systems, natural conditions are probably different, because only low amounts of virus or better virus-infected cells from a donor might have mediated heterotransmission. Such a situation has not been modeled in detail, but it can be predicted that under such conditions the virus becomes efficiently expressed only rarely. Therefore, there might be a bridge connecting trans-class and trans-species retrovirus heterotransmission.

Some experimental data obtained from long-term follow-up of retrovirus persistence (Trejbalová et al., 1999) suggest that provirus silencing discussed in the first and also next chapter might play a role in the control of low-dose infection and long-term retrovirus persistence.

4. Provirus silencing and expression in differentiated cells

In order to investigate the role of cell differentiation in relation to retrovirus infection, several models have been designed, but most thoroughly were analyzed EC cells, which differentiate in vitro. The entrée to this problem was given by the group of Jaenisch (Stewart et al., 1982).

who discovered that MLV integrated in EC cells becomes methylated and unexpressed. In a series of papers from different laboratories (reviewed by Challita et al., 1995) it was established that in order to ensure MLV expression in EC, introduction of a transcription factor Sp1 binding site into LTR is required together with inactivation of the negative control region (NCR) in LTR as well as of the region of the primer binding site. As the third negatively acting element, one out of two direct repeats in LTR was recognized (Hawley et al., 1994). There is still room left for MLV improved expression, as exemplified by insertion into LTR of an antimethylation fragment from the region upstream of the Thy1 gene (Challita et al., 1995).

All the above-mentioned modifications act synergistically. Recently, MLV LTR lacking known negative elements due to deletion spanning most of the 5′ end LTR portion has been constructed (Osborne et al., 1999). As a result of such a deletion, about half of 13 CpG sites within LTR were also lost. Such truncated LTR, essentially stripped of enhancer elements, was employed for generation of a retroviral vector harboring the neomycin resistance (*neo*) reporter gene equipped with internal β-globin promoter, which after infection of EC cells displayed expression in 70% of cells, the highest efficiency obtained so far.

We have not yet reached the end of the journey to optimal retroviral vector function in differentiated cells. It is still possible that additional, so far undefined sequences should be inactivated, altered or inserted. Of special importance might be CpG dinucleotides present in LTR, especially at the start of transcription. Elements acting as antimethylation signals, interfering with silencers or ensuring position-independent gene expression, such as locus control regions (LCR), should be tested for their ability to ensure retrovirus or retroviral vector expression in differentiated cells. Recently, the chicken hypersensitive site 4 (cHS4) of the chicken globin LCR, acting as an insulator, when cloned into MLV LTR was shown to increase the probability of integrated proviruses expression and to decrease the level of de novo methylation of the 5′LTR in murine erythroleukemia (MEL) cells (Rivella et al., 2000). In addition, the human β interferon scaffold attachment region (IFN-SAR), when inserted in retroviral LTR, prevented its methylation and ensured vector expression in a stably transfected line of human T cells. The vector expression has been kept for several months and included also multiple proviral copies (Agarwal et al., 1998; Dang et al., 2000). We can therefore stress the point that the problem of permissiveness of differentiated cells to retrovirus infection goes far beyond EC cells and that other differentiated cells, such as hematopoietic or hepatic cells, should probably require not only prevention of vector downregulation, but even some more specific cell changes like activation of steps triggering the cell cycle (rev. Emerman, 2000).

Dealing with provirus silencing in differentiated cells we focused on provirus methylation as an epigenetic DNA modification described repeatedly in conjunction with provirus downregulation. However, it is not clear so far whether provirus methylation acts as the primary cause or whether it only conserves transcriptional repression. In provirus methylation studies, usually a general increase in CpG methylation has been measured. However, some more precise data have been obtained showing that methylation of one particular CpG has a decisive effect. This is the case of HIV, where methylation of only one CpG at position −143 (in the vicinity of NF-κB and Sp1 binding sites) in HIV LTR results in 70% inhibition of the reporter expression (Bednarik et al., 1990; Schulze-Forster et al., 1990). Similarly, single CpG site methylation in the RSV LTR U5 region nearby the transcription start leads to a clear decrease in provirus expression (Hejnar et al., 1999). As discussed later, the density of CpGs is also of importance.

It has been recognized on other gene models that methylation provides a signal for association with methyl-CpG-binding protein 2 (MeCP2), which, through the adaptor protein Sin3A, recruits the histone deacetylase (HDAC) (rev. Razin, 1998; Ng and Bird, 1999; Knoepfler and Eisenman, 1999). A similar situation was disclosed in the case of MeCP1, which produces a complex composed of the MBD protein containing the methyl-CpG-binding domain and two members of the HDAC family (rev. Bird and Wolffe, 1999). Thus, there is a proven link between CpG methylation and chromatin deacetylation.

How are these findings related to provirus silencing? It should be noted that MeCP1 has been already shown to bind to methylated LTR of myeloproliferative sarcoma virus, suppressing its LTR activity (Boyes and Bird, 1991). This problem has been recently approached by Lorincz et al. (2000). They employed the MLV LTR-driven green fluorescence protein (GFP) gene and followed GFP expression after infection of MEL cells with this vector. Cell clones that displayed silenced GFP were isolated and it was disclosed that early after silencing the proviruses became methylated to a low degree. Such clones could have been reactivated by trichostatin (TSA), which inhibits HDAC-MeCP2 complexes. During prolonged cell cultivation the vector has been increasingly methylated and in the hypermethylated state provirus expression was induced only by combination of TSA and 5-azacytidine (5-azaC), which acts as a demethylation agent. These results suggest that provirus downregulation is a dynamic process and that the possibility of its reactivation depends upon the density of methylation.

All the data obtained so far point to the important role of methylation in provirus silencing in general, and therefore strategies preventing methylation of retroviral vectors as well as putative blocking of methyl-CpG-binging proteins should contribute to more efficient gene therapy applied to differentiating or differentiated cells.

There is, in addition, an important question: whether and why retroviruses are more efficiently recognized and silenced than any foreign DNA introduced in a genome. In other words, does there exist a cell genome surveillance mechanism (analogous but not homologous to immunity)

that ensures downregulation of retroviral sequences? Methylation has been already proposed to fulfil such a duty (Doerfler, 1991; Yoder et al., 1997). However, some signals should be involved that attract methyltransferase to an integrated provirus, or even to a specific DNA structure common to viral integration intermediates (Bestor, 1987). Such signals might be provided by flanking chromosomal sequences or by the proviral structure itself. Especially LTR could be recognized as unusual direct repeats. Similar structures in lower eukaryotes trigger gene silencing (rev. Wolffe and Matzke, 1999). Furthermore, MLV LTR binding zinc finger transcription factor YY-1 (Flanagan et al., 1992) was shown to represent a homolog of Sin3 (Yang et al., 1996), which has been already identified as a protein complexing with HDAC involved in formation of transcriptionally inactive chromatine. It should be also taken into account that in some heterologous and differentiated cells, proteins activating LTR might be underrepresented or that such cells produce altered isomorphic proteins, which could be inactive and/or could interfere with factors required for LTR activation. Therefore, retroviral genomes should be screened also from the point of view of sequences and DNA-protein complexes which might contribute to provirus silencing, of course, in context with flanking chromosomal DNA. New techniques such as inverse polymerase chain reaction should facilitate this demanding task.

Acknowledgements

The authors highly acknowledge the excellent editorial help of Šárka Takáčová. These studies, which involve several interconnected areas of research, were supported by Grant No. 1R03 TW00 155-01A1 (Fogarty International Research Collaboration Award), Grants No. 301/94/0713, No. 203/96/0883, No. 312/96/K205, No. 312/98/0825 and No. 312/97/P082 from the Grant Agency of the Czech Republic.

References

Agarwal, M., Austin, T.W., Morel, F., Chen, J., Böhnlein, E., Plavec, I., 1998. Scaffold attachment region-mediated enhancement of retroviral vector expression in primary T cells. J. Virol. 72, 3720–3728.

Altanerova, V., Ban, J., Altaner, C., 1989. Induction of immune deficiency syndrome in rabbits by bovine leukaemia virus. AIDS 3, 755–758.

Anderssen, S., Sjøttem, E., Svineng, G., Johansen, T., 1997. Comparative analysis of LTR of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset African green monkey, and man. Virology 234, 14–30.

Bednarik, D.P., Cook, J.A., Pitha, P.M., 1990. Inactivation of the HIV LTR by DNA CpG methylation Evidence for a role in latency. EMBO J. 9, 1157–1164.

Bestor, T.H., 1987. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. Nucleic Acids Res. 15, 3835–3843.

Bird, A.P., Wolffe, A.P., 1999. Methylation-induced repression - belts, braces, and chromatin. Cell 99, 451–454.

Boyes, J., Bird, A., 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 64, 1123–1134.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., Cedar, H., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438.

Burny, A., Bruck, C., Cleuter, Y., Couez, D., Deschamps, J., Ghysdael, J., Gregoire, D., Kettmann, R., Mammerickx, M., Marbaix, G., 1985. Bovine leukaemia virus, a versatile agent with various pathogenic effects in various animal species. Cancer Res. 45, 4578s–4582s.

Challita, P.-M., Skelton, D., El-Khoueiry, A., Yu, X.-J., Weinberg, K., Kohn, D.B., 1995. Multiple modifications in *cis* elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. J. Virol. 69, 748–755.

Chen, Z., Telfer, P., Gettie, A., Reed, P., Zhang, L., Ho, D.D., Marx, P.A., 1996. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. J. Virol. 70, 3617–3627.

Dang, Q., Auten, J., Plavec, I., 2000. Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. J. Virol. 74, 2671–2678.

Doerfler, W., 1991. Patterns of DNA methylation - evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. Biol. Chem. Hoppe-Seyler 372, 557–564.

Duran-Reynals, F., 1942. The reciprocal infection of ducks and chickens with tumor inducing viruses. Cancer Res. 2, 343–369.

Emerman, M., 2000. Learning from lentiviruses. Nature Genet. 24, 8–9.

Fauci, A.S., 1996. Host factors and the pathogenesis of HIV-induced disease. Nature 384, 529–534.

Faxvaag, A., Dai, H.Y., Aarseth, H., Dalen, A.B., 1993. A low oncogenic variant of friend murine leukemia virus with strong immunosuppressive properties. Arch. Virol. 131, 265–275.

Fincham, V.J., Wyke, J.A., 1991. Differences between cellular integration sites of transcribed and nontranscribed Rous sarcoma proviruses. J. Virol. 65, 461–463.

Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B.Z., Appella, E., Ozato, K., 1992. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol. Cell. Biol. 12, 38–44.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Hahn, B.H., 1992. Human infection by genetically diverse SIV_{SM}-related HIV-2 in West Africa. Nature 358, 495–499.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.M., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee. *Pan troglodytes troglodytes*. Nature 397, 436–441.

Hawley, R.G., Lieu, F.H., Fong, A.Z., Hawley, T.S., 1994. Versatile retroviral vectors for potential use in gene therapy. Gene Ther. 1, 136–138.

Hejnar, J., Plachý, J., Geryk, J., Machoň, O., Trejbalová, K., Guntaka, R.V., Svoboda, J., 1999. Inhibition of the Rous sarcoma virus long terminal repeat-driven transcription by in vitro methylation: different sensitivity in permissive chicken cells versus mammalian cells. Virology 255, 171–181.

Hirsch, V.M., Olmsted, R.A., Murphy-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIV_{sm}) closely related to HIV-2. Nature 339, 389–392.

Karakoz, I., Geryk, J., Svoboda, J., 1980. In vivo effect of three transformation-defective mutants of subgroup C avian sarcoma viruses. In: Zadrazil, S., Sponar, J. (Eds.), DNA – Recombination. Interactions and Repair. Pergamon Press, Oxford and New York, pp. 435–440.

Knoepfler, P.S., Eisenman, R.N., 1999. Sin meets NuRD and other tails of repression. Cell 99, 447–450.

Kucerova, L., Altanerova, V., Altaner, C., Boris-Lawrie, K., 1999. Bovine

leukemia virus structural gene vectors are immunogenic and lack pathogenicity in a rabbit model. J. Virol. 73, 8160–8166.

Levy, J.A., 1977. Murine xenotropic type C viruses III. Phenotypic mixing with avian leukosis and sarcoma viruses. Virology 77, 811–825.

Levy, J.A., 1978. Xenotropic type C viruses. Curr. Top. Microbiol. Immunol. 79, 111–213.

Levy, J.A., Oleszko, O., Dimpfl, J., Lau, D., Rigdon, R.H., Jones, J., Avery, R., 1982. Murine xenotropic type C viruses IV. Replication and pathogenesis in ducks. J. Gen. Virol. 61, 65–74.

Lorincz, M.C., Schübeler, D., Goeke, S.C., Walters, M., Groudine, M., Martin, D.I.K., 2000. Dynamic analysis of proviral induction and de novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. Mol. Cell. Biol. 20, 842–850.

Machoň, O., Strmen, V., Hejnar, J., Geryk, J., Svoboda, J., 1998. Sp1 binding sites inserted into the Rous sarcoma virus long terminal repeat enhance LTR-driven gene expression. Gene 208, 73–82.

Macleod, D., Charlton, J., Mullins, J., Bird, A.P., 1994. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292.

Mazgareanu, S., Muller, J.G., Czub, S., Schimmer, S., Bredt, M., Czub, M., 1998. Suppression of rat bone marrow cells by Friend murine leukemia virus envelope proteins. Virology 242, 357–365.

Mummaneni, P., Bishop, P.L., Turker, M.S., 1993. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. J. Biol. Chem. 268, 552–558.

Nelson, D.T., Goodchild, N.L., Mager, D.L., 1996. Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. Virology 220, 2213–2218.

Ng, H.-H., Bird, A., 1999. DNA methylation and chromatin modification. Curr. Opin. Genet. Dev. 9, 158–163.

Osborne, C.S., Pasceri, P., Singal, R., Sukonnik, T., Ginder, G.D., Ellis, J., 1999. Amelioration of retroviral vector silencing in locus control region β-globin-transgenic mice and transduced F9 embryonic cells. J. Virol. 73, 5490–5496.

Prince, V.E., Rigby, P.W.J., 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65, 1803–1811.

Razin, A., 1998. CpG methylation, chromatin structure and gene silencing – a three-way connection. EMBO J. 17, 4905–4908.

Rivella, S., Callegari, J.A., May, C., Tan, C.W., Sadelain, M., 2000. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. J. Virol. 74, 4679–4687.

Rynditch, A., Kadi, R., Geryk, J., Zoubak, S., Svoboda, J., Bernardi, G., 1991. The isopycnic compartmentalized integration of Rous sacoma virus sequences. Gene 106, 165–172.

Saha, K., Yuen, P.H., Wong, P.K.Y., 1994. Murine retrovirus-induced depletion of T cells is mediated through activation-induced death by apoptosis. J. Virol. 68, 2735–2740.

Schulze-Forster, K., Götz, F., Wagner, H., Kroger, H., Simon, D., 1990. Transcription of HIV 1 is inhibited by DNA methylation. Biochem. Biophys. Res. Commun. 168, 141–147.

Sharp, P.M., Robertson, D.L., Hahn, B.H., 1995. Cross-species transmission and recombination of 'AIDS' viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 349, 41–47.

Sjøttem, E., Anderssen, S., Johansen, T., 1996. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. J. Virol. 70, 188–198.

Smith, R.E., Schmidt, E.V., 1982. Induction of anemia by avian leukosis viruses of five subgroups. Virology 117, 516–518.

Stepanets, V., Vernerová, Z., Vilhelmová, M., Geryk, J., Hejnar, J., Weichold, F.F., Svoboda, J., 2000. Intraembryonic avian leukosis virus subgroup C (ALV-C) inoculation producing wasting disease in ducks soon after hatching (submitted).

Stewart, C.L., Stuhlmann, H., Jahner, D., Jaenisch, R., 1982. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79, 4098–4102.

Svoboda, J., 1986. Rous sarcoma virus. Intervirology 26, 1–60.

Svoboda, J., 1998. Molecular biology of cell non-permissiveness to retroviruses. Has the time come? Gene 206, 153–163.

Trejbalová, K., Geblhard, K., Vernerová, Z., Dušek, L., Geryk, J., Hejnar, J., Haase, A.T., Svoboda, J., 1999. Proviral load and expression of avian leukosis viruses subgroup C in long-term persistently infected heterologous hosts (ducks). Arch. Virol. 144, 1779–1807.

Weiss, R.A. Wrangham, R.W., 1999. From *pan* to pandemic. Nature 385–386.

Wolffe, A.P., Matzke, M.A., 1999. Epigenetics: regulation through repression. Science 286, 481–486.

Wyatt, C.R., Wingett, D., White, J.S., Buck, C.D., Knowles, D., Reeves, R., Magnuson, N.S., 1989. Persistent infection of rabbits with bovine leukemia virus associated with development of immune dysfunction. J. Virol., 63, 4498–4506.

Yang, W.M., Inouye, C., Zeng, Y., Bearss, D., Seto, E., 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc. Natl. Acad. Sci. USA 93, 12845–12850.

Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335–340.

Young, W.-B., Lindberg, G.L., Link Jr, C.J., 2000. DNA methylation of helper virus increases genetic instability of retroviral producer cells. J. Virol. 74, 3177–3187.

TAB 7

Clin Genet 2001: 59: 17–24
Printed in Ireland. All rights reserved

CLINICAL GENETICS
ISSN 0009-9163

Developmental Biology: Frontiers for Clinical Genetics

Section Editor: Roderick R McInnes
e-mail: mcinnes@sickkids.on.ca

The beta-globin locus control region versus gene therapy vectors: a struggle for expression

Ellis J, Pannell D. The beta-globin locus control region versus gene therapy vectors: a struggle for expression.
Clin Genet 2001: 59: 17–24. © Munksgaard, 2001

Developmental control of gene expression has a major impact on the design of β-globin retrovirus vectors for hematopoietic stem cell gene therapy of β-thalassemia. It is obvious that the endogenous locus control region (LCR) elements that drive β-globin gene expression in transgenic mice must be included in these vectors. However, the specific elements to use are not clear and require an understanding of LCR action. Moreover, retrovirus vectors contain silencer elements that function in stem cells and are dominant to LCR function. Recent studies on LCRβ-globin transgenes and retrovirus silencing suggest ways to overcome this silencing effect after transfer into stem cells and carefully designed lentivirus vectors have exciting therapeutic benefit in animal models of β-thalassemia. By building on 15 years of development, LCRβ-globin vectors are now being tested in preclinical animal models and may ultimately lead to the long-sought cure for this genetic disease.

J Ellis and D Pannell

Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON, and Department of Molecular and Medical Genetics, University of Toronto, Canada.

Key words: chromatin – gene silencing – gene therapy – LCR

Corresponding author: J Ellis, Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8. Tel: +416 813 7295; fax: +416 813 8883; e-mail: jellis@sickkids.on.ca

Received 12 October 2000, revised and accepted for publication 16 October 2000

β-thalassemia is caused by genetic defects that reduce β-globin protein levels (1). The resulting imbalance of β-globin to α-globin chains results in a severe anemia that is generally treated by repeated blood transfusions. A consequence of such regular blood transfusion is the increased risk of exposure to virus-infected blood supplies, and eventually leads to iron overload that can cause organ failure (2). Present iron chelation treatments are controversial (3), and the only cure available is bone marrow transplantation from a matched sibling if available. An attractive alternative is to perform gene therapy to deliver a human β-globin gene into hematopoietic stem cells (HSC) from the patient (4, 5).

For β-globin gene therapy to be successful, it is essential that the transferred gene be expressed to the correct level. This goal requires a detailed knowledge of the mechanism and *cis*-acting sequences that control β-globin expression during development. Surprisingly, inclusion of the appropriate regulatory elements may not be sufficient to obtain therapeutic levels during gene therapy as the vectors used to deliver the gene are frequently silenced in transduced stem cells. Hence, a better understanding of the mechanism of vector silencing in stem cells is also required. Here, we review the regulatory elements that control β-globin gene expression during development and their use in β-globin retrovirus vectors, outline the evidence that retrovirus and lentivirus vectors are silenced in stem cells and potential means to overcome this silencing, and we conclude with preclinical animal models to test promising β-globin gene therapy vectors.

β-globin gene expression during development

The β-globin gene is part of a cluster of highly related globin genes located on Chr 11p15 in humans (6). These genes are arranged in the same order as they are expressed during development (Fig. 1). The ε-globin gene is expressed in the blood islands of the yolk sac, the site of hemato-

poiesis then switches to the fetal liver where the γ-globin genes are expressed, and shortly after birth hematopoiesis switches to the bone marrow where the δ- and β-globin genes are expressed to very low and high levels, respectively. Many mutations that cause β-thalassemia have been described and some have been informative with regard to the sequences and molecular mechanisms that control globin gene switching. For example, it is clear that point mutations in the γ-globin promoters can enhance expression of these genes in adults causing hereditary persistence of fetal hemoglobin (HPFH). Deletion of upstream sequences that include the locus control region (LCR) results in a lack of expression from the still intact globin genes. These data demonstrate that promoter sequences and the LCR are important control elements for globin expression.

The β-globin locus control region

The LCR is composed of at least four DNaseI hypersensitive sites (HS) located upstream of the locus (Fig. 1) (7–9). The presence of HS indicates that *trans*-acting factors are binding to these regions and displacing or destabilizing nucleosomes. Nucleosomes are the basic units of chromatin and condense DNA around an octomer of the histone proteins H2A, H2B, H3 and H4 (10). Expressed genes are located an 'open' chromatin that is more accessible to *trans*-acting factors and in general contain nucleosomes with highly acetylated histones. In contrast, 'closed' chromatin generally has deacetylated histones and is bound by the linker histone H1, is less accessible to DNA binding factors and genes in these regions are not expressed. Chromatin structure is modulated using many chromatin remodelling complexes (11, 12).

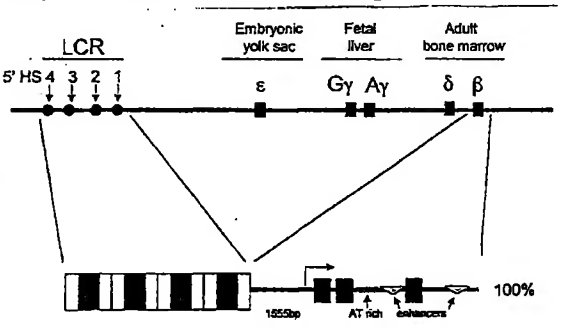

Fig. 1. Gene structure and expression pattern of the human β-globin locus. The locus control region (LCR) is composed of at least 4 DNaseI hypersensitive sites (HS) located upstream of the cluster. *Trans*-acting factors (red circles) bind to the HS. In transgenic mice, LCR activity directs copy number-dependent, position-independent transgene expression. Full levels of β-globin transgene expression are obtained in the presence of all four HS including their core elements (red boxes).

The mechanism by which the LCR controls β-globin gene expression has been extensively studied, primarily using transgenic and knockout mice (6). It is important to realize that these two assays manipulate the genes in quite different ways and the results are not always complementary or in agreement with each other. Transgenic mice contain the human β-globin gene transferred into novel or ectopic integration sites; whereas, knockout mice manipulate the endogenous native locus in the mouse. In transgenic mice, human β-globin transgenes are silent at most integration sites or transcribed to about 1% of the endogenous mouse βmajor level. In contrast, addition of the LCR including all four HS to the β-globin transgene results in expression to about 100% levels at all integration sites, and expression is copy number-dependent (Fig. 1) (7, 13, 14). This copy number-dependent, position-independent transgene expression is unusual and is the defining feature of LCR activity. Further investigation demonstrated that individual HS2, HS3 and HS4 elements and their smaller 'cores' of approximately 200–300 bp, also direct copy number-dependent transgene expression but to lower levels (10–25%) (15–20).

The LCR is often referred to as an enhancer, but does not have classic enhancer activity because it does not function equally well in either orientation (21). Rather, it appears that complete LCR activity requires all four HS (22), and these have some distinct roles. For example, HS3 can activate β-globin transgenes at all single copy integration sites where it establishes open chromatin and remodels chromatin on the promoter to permit expression (13). In contrast, although HS2 has strong enhancer activity in transient transfection studies (23), it is unable to direct expression in single copy transgenic mice (24). These data suggest that at ectopic sites, the HS function together as a unit, making the LCR sufficient to open chromatin and enhance full expression of β-globin transgenes.

Open chromatin is likely to be established by the binding of erythroid *trans*-acting factors that recruit chromatin remodelling complexes (25–31), as has been described for histone acetylation changes on active human β-globin genes (32). This open chromatin may not extend throughout transgenes containing the entire human LCRβ-globin cluster, as different domains that correlate with the presence of low level intergenic transcription have been described during globin switching in mice (33).

Two models of LCR activity

The transgenic mouse data have largely been interpreted as supporting a holocomplex model of LCR

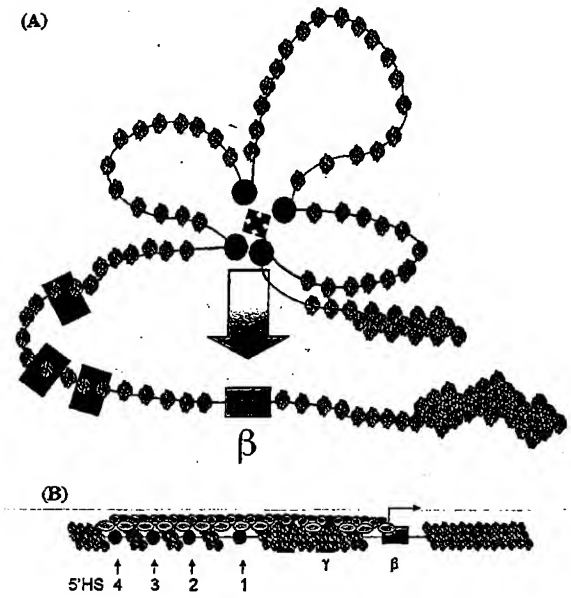

Fig. 2. Two models of LCR activity. A) The holocomplex model suggests that factors (red circles) bound to the HS interact with each other by DNA looping (arrows) to form a single LCR holocomplex that loops to activate expression from the appropriate globin promoter. B) The linking model suggests that the LCR serves to ensure that factors (coloured circles) are bound throughout the locus. Nucleosomes in 'open' chromatin are represented as dispersed green circles, 'closed' chromatin as condensed solenoids of green circles.

action at ectopic sites (34). Once open chromatin has been established, it is proposed that each of the HS then interacts with each other by DNA looping mediated via the bound factors to form an LCR holocomplex (Fig. 2A). The holocomplex would then interact with a single globin gene in the cluster, and switching during development would be accomplished by stage-specific silencer elements associated with the ε- and γ-globin genes. Although there is no direct physical evidence for DNA looping between the LCR and the globin promoters *in vivo*, the holocomplex model is supported by findings that only a single globin gene is transcribed at a time in transgenic mice containing the whole human LCRβ-globin cluster (35), and that the LCR preferentially activates genes closest to it (36). Deletions that remove only a single HS 'core' element drastically reduce transgene expression in comparison to those that delete an entire HS fragment (37, 38), indicating that removal of a core creates a defective holocomplex (39).

Quite different conclusions have been arrived at using knockout technology on the mouse β-globin locus (40). Deletions of individual or all the HS in the endogenous locus do not alter chromatin structure and have relatively minor effects on expression of the globin genes (41, 42). These data suggest that the LCR is not required for chromatin opening at the endogenous mouse β-globin locus, and suggest that more distant elements control chromatin structure (43). A linking model that does not invoke DNA looping has been proposed to explain the knockout results (Fig. 2B) (44). In this model, the function of the LCR is to enhance β-globin expression by ensuring that factors are bound at intervals across the cluster and that the gene is localized to the right nuclear compartment (32). The linking model is not consistent with the ability of the LCR to open chromatin at ectopic transgene sites, but the holocomplex model cannot easily explain the effect of LCR deletions in the mouse β-globin locus. As described below, the two models are not necessarily mutually exclusive and may be strengthened by being merged. For the purpose of gene therapy where globin expression cassettes delivered by viral vectors must express at ectopic sites, it will be important to design the cassettes based on transgene constructs that express to high levels at single copy integration sites.

LCRβ-globin expression cassettes for gene therapy

To express therapeutic levels of β-globin from gene therapy cassettes, full expression levels should be obtained from a single copy integration in order to convert a null thalassemia into an asymptomatic carrier state. Initial β-globin gene therapy cassettes were designed to be as small as possible to facilitate gene transfer and used either cDNA or genomic β-globin fragments controlled by minimal promoters (45–47). Addition of small HS core elements improves expression in tissue culture experiments, but were largely disappointing when transferred into mouse bone marrow cells (48, 49). Recent single copy transgenic mouse experiments now demonstrate that full expression by the LCR requires all four HS and specific elements within the β-globin gene including the −1555 bp promoter and the 3′ enhancer (Fig. 3 top) (50). As this 8.8-kb cassette is too large for conventional retrovirus vectors, smaller constructs that express highly are required for retrovirus delivery. A very promising new 3.9-kb cassette expresses γ-globin mRNA to 70% levels in single copy transgenic mice, using β-globin promoter, intron 2 and 3′ enhancer elements (Fig. 3 centre) (51). This construct, or others shown to function effectively in transgenic mice, may ultimately prove to be best suited for gene therapy and take advantage of the anti-sickling properties of γ-globin. In addition, these data demonstrate that the LCR must functionally interact with more than just the promoter. The simplest

interpretation merges the holocomplex and linking models by suggesting that the LCR loops to interact with the promoter, but factors must also be bound throughout the gene.

Retrovirus and lentivirus vectors

Retrovirus vectors have been the method of choice for delivering β-globin genes into hematopoietic stem cells because they stably integrate at single copy into the genome (52). In practice, it has been very difficult to obtain high titer β-globin retrovirus vectors due to instability of the LCR elements and globin intron 2 sequences (48, 49). To stabilize transmission of intact β-globin genes it is necessary to use only certain combinations of HS sites and to delete an AT-rich region in intron 2. Although these modifications permit generation of high titer retrovirus, it is now apparent that the deleted AT-rich sequence is required for high level expression (51).

A limitation of retrovirus vectors is that they integrate only into cycling cells, and the target HSC for β-globin gene therapy tend to be non-cycling. In this regard, lentivirus vectors based on HIV-1 (53) are far superior as they integrate into non-cycling cells and contain the RRE element that is bound by the Rev protein to stabilize the virus genomic RNA (52). One exciting report of a β-globin lentivirus vector demonstrates that it is possible to transmit large LCR fragments coupled to a β-globin cassette with a small promoter, AT-rich deleted intron 2 and no 3′ enhancer (Fig. 3 bottom). Therapeutic levels of β-globin mRNA and protein were shown in transduced bone marrow in a mouse model of β-thalassemia (54).

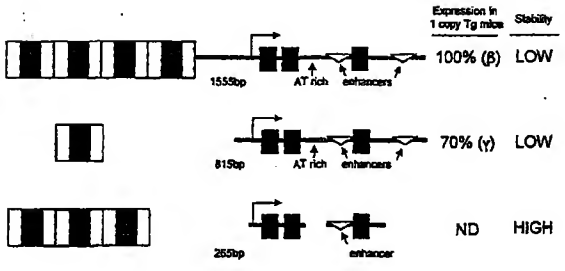

Fig. 3. LCRβ-globin expression cassettes for gene therapy. Full expression in single copy transgenic mice requires an 8.8 kb construct (top) including all four HS, the large β-globin promoter, β-globin intron 2 AT-rich region, and the intron 2 and 3′ enhancers (yellow triangles). High level expression of γ-globin exons (Blue boxes) is obtained from a smaller 3.9 kb β/γ-globin hybrid cassette (centre). As both of these cassettes are not stable in retrovirus vectors, cassettes with improved stability (bottom) have deletions of deleterious sequences.

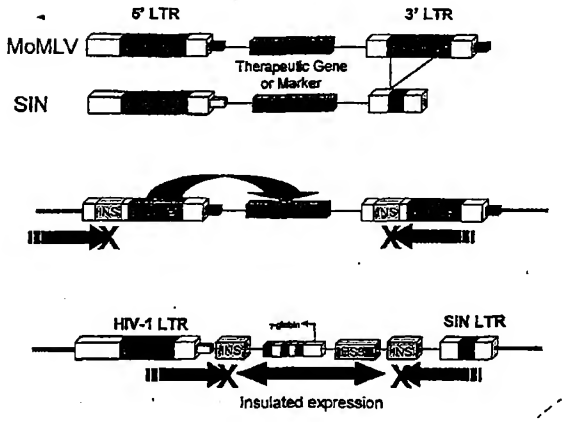

Fig. 4. Retrovirus silencing in stem cells. Silencer elements in MoMLV vectors are primarily located in the LTRs (top), and silencing can be ameliorated but not eliminated by SIN mutations in the 3′LTR. Insulator elements (INS) in the LTR (centre) block silencing by position effects (red arrows with X) in MEL cells but not retrovirus silencing in stem cells (large red arrow). The ideal vector (bottom) uses a SIN lentivirus vector, insulator elements flanking a small HS3β/γ-globin expression cassette to prevent silencing (green arrow), and no marker gene.

Silencing of retrovirus vectors in stem cells

Once the β-globin expression cassette has been transduced into an HSC, it must express to appropriate levels in erythroid cells produced from the HSC. One concern that has severely limited clinical gene therapy is that retrovirus-transduced genes are often silenced in HSC (52, 55–57). It has long been known that retroviruses are silenced in embryonic stem cells and in transgenic mice (58, 59). Moreover, it has been shown that retrovirus and lentivirus sequences silence LCRβ-globin transgenes in mice (60, 61), and that the transgenes fail to open chromatin (60). These data demonstrate that the vectors have silencer elements, and suggest that they are recognized by specific factors that are restricted to stem cells in mammals.

There are several possible means to overcome vector silencing in stem cells. First, the silencer elements can be defined in the retrovirus or lentivirus vectors and, if not required for virus replication, can be removed. As the silencers tend to be located in the viral long terminal repeats (LTRs) that control virus transcription (Fig. 4 top), mutations produce self-inactivating (SIN) virus that is inherently safer. Several SIN retrovirus vectors have been developed from Moloney murine leukemia virus that express to higher levels in ES cells and transgenic mice (62–65). Nevertheless, none of the SIN vectors express at all integration sites in stem cells. To overcome this residual silencing in the SIN vectors, it may be possible to

incorporate insulator elements to block the spread of silencing from surrounding retrovirus sequences (66). The chicken β-globin cHS4 element is an insulator (66, 67) and has been placed in the LTR of retrovirus vectors to block position effects in murine erythroleukemia (MEL) cells (Fig. 4 centre) (68, 69). However, an insulator in the LTR does not shield internal genes from silencing established on the retrovirus sequences in ES cells (68). A better construct design would position insulators on both sides of the internal gene rather than in the LTR (Fig. 4 centre).

Mechanism of retrovirus vector silencing

An understanding of the mechanism of retrovirus silencing may permit interventions that prevent its establishment or maintenance (70, 71). Many groups have correlated *de novo* cytosine methylation (72) of CpG dinucleotides in retroviruses with the silenced state (58, 59, 68, 73, 74), and expression can be reactivated to a low level using the methylation inhibitor 5 AzaCytidine (5AzaC) (75). However, this indirect evidence does not address whether methylation is a cause of silencing or a consequence. Recent direct evidence that retrovirus silencing is independent of *de novo* methylase function has been obtained using dnmt3 knockout ES cells and transgenic *Drosophila* that have no methylase activity (60). As chromatin of retrovirus-silenced LCRβ-globin transgenes in mice is inaccessible to DNaseI and marked by deacetylated histone H3 and bound linker H1 (60), it appears that chromatin modifications play a role in retrovirus silencing. Attempts to relieve silencing may, therefore, require inhibitors of histone deacetylases (HDAC) or of H1 binding.

Silencing is often established by one pathway and maintained by another. Time course experiments in ES cells demonstrate that most retroviruses are silenced within 2 days but some integration sites escape complete silencing and express to low levels (60, 76, 77). As methylation is not detectable by 2 days post-infection, methylation is likely to be a consequence of, or secondary step in, retrovirus silencing. Most experiments have focussed on the subset of infected cells that initially express. Over time, these are gradually silenced in a process known as extinction. In infected MEL cells, extinction can be reversed early in the process using the HDAC inhibitor Trichostatin A (TSA) (78). However, the methylation inhibitor 5 AzaC is required in addition to TSA to overcome extinction at later time points. These data suggest that methylation is an important secondary or associated step in extinction of virus expression in mature cell types. The ability of TSA or 5 AzaC to activate expression in the majority of transduced cells that are completely silenced from the outset has not been rigourously tested to date. In summary, efforts to prevent retrovirus silencing using methylation and HDAC inhibitors hold promise, but require more knowledge of the mechanism and demonstration of their utility in silenced stem cell populations prior to extinction.

An ideal LCRβ-globin lentivirus vector

The ideal β-globin gene therapy vector should stably integrate into an HSC at high efficiency and be expressed to near endogenous levels at single copy. To accomplish this goal, lentivirus vectors have a clear advantage in their ability to infect non-cycling stem cells and stably transmit large LCRβ-globin expression cassettes. The best existing LCRβ-globin lentivirus vector has these features (54), with the additional advantage of omitting a selectable marker gene. Although marker genes are convenient for determining transduction frequencies, most are derived from non-mammalian sources and may themselves be subject to gene silencing effects. However, the vector can be optimized further to direct expression at all integration sites and for vector safety (Fig. 4 bottom). First, a SIN version of the lentiviral vectors must now be used with third generation packaging systems designed to prevent recombination events that generate replication-competent HIV-1 virus (79). The SIN lentivirus vector will not only prevent rescue and spread of the vector by any helper virus, but may also improve expression of the LCRβ-globin cassette (60). Second, the LCRβ-globin cassette should be flanked by insulator elements that are known to block silencing. In practise, this may require a different insulator on one side than the other to avoid recombination events that delete the LCRβ-globin cassette. Finally, the LCRβ-globin cassette should express highly at single copy in transgenic mice. Such constructs use HS3 coupled to a large β-globin promoter, the β-globin intron 2 including both the AT-rich region and enhancer and the 3′ enhancer (51). Use of hybrid genes permits expression of anti-sickling γ- or δ-globin coding sequences instead of β-globin exons (51, 80). A combination of these components should create a lentivirus vector that is safe and expresses therapeutic levels of globin.

Preclinical models of β-thalassemia

Promising β-globin gene therapy vectors have been tested primarily in MEL cells, or in infected mouse

bone marrow. Expression in erythroid cells derived from primitive progenitor cells can be assayed using CFU-S assays in which spleen colonies are formed after 12 days *in vivo*. Expression in cells derived from an infected HSC must be assayed after long-term repopulation assays, followed by secondary transplantation. To show therapeutic efficacy, these long-term studies should be performed first on mouse models of β-thalassemia and sickle cell anemia. A variety of these models have been created by gene targeting and transgenic technology (81–83). These mouse models can be corrected by expression of γ-globin transgenes, and it has been shown that an LCRβ-globin lentivirus vector can express β-globin mRNA and protein in these models with therapeutic benefit (54).

Despite success with the mouse models, it has proved much more difficult to transduce human HSC than murine HSC (84). Fortunately, another preclinical model of β-globin gene therapy into human stem cells is available (Fig. 5). The NOD-Scid mouse is severely immunocompromized and fails to reject human bone marrow transplants (85). Transplanted human HSC home to the mouse bone marrow where they are supported by the hematopoietic microenvironment. Human bone marrow cells from β-thalassemia and sickle cell patients have been shown to repopulate the bone marrow of these mice and generate human red cells that mimic the disease. This system is very well suited to test expression from LCRβ-globin lentivirus vectors in human stem cells. To this end, patient bone marrow or sorted HSC would be infected with the lentivector *in vitro* prior to *in vivo* assay in the NOD/Scid mice. Hence, long-term expression in human stem cells can be assayed *in vivo* without exposing patients to experimental lentivirus vectors.

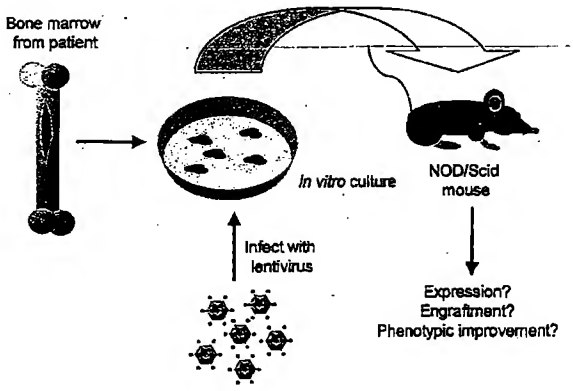

Fig. 5. The NOD-Scid preclinical mouse model of human stem cell gene therapy for β-thalassemia or sickle cell anemia. Patient bone marrow stem cells are infected with the β-globin lentivirus vector *in vitro*, prior to *in vivo* engraftment into immunocompromised NOD-Scid mice. Gene transfer into engrafted human stem cells, β-globin transgene expression, and phenotypic improvement can then be monitored without exposing patients to experimental lentivirus vectors.

Future prospects

The first retrovirus vectors for β-globin gene therapy were designed over 15 years ago, and through slow careful research many obstacles were discovered and gradually surmounted. The success of a well-designed LCRβ-globin lentivirus vector in correcting a mouse β-thalassemia model is a milestone in this process that can now be completed through incremental improvements to vector expression and safety. Ultimately, validation of β-globin gene therapy in human stem cells using the NOD-Scid preclinical model will justify clinical trials of this exciting potential cure for hemoglobinopathies.

Acknowledgements

We acknowledge the Medical Research Council (MRC) of Canada grant support (to JE) for our β-globin expression and retrovirus silencing research and Hospital for Sick Children Foundation, OGS and MRC Doctoral Research Awards to DP. We regret that many fine studies relevant to this review have been omitted due to space constraints.

References

1. Weatherall DJ. The Thalassemias. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 183–226.
2. Olivieri NF. The beta-thalassemias. N Engl J Med 1999: 341 (2): 99–109.
3. Olivieri NF et al. Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major. N Engl J Med 1998: 339 (7): 417–423.
4. Persons DA, Nienhuis AW. Gene therapy for the hemoglobin disorders: past, present, and future. Proc Natl Acad Sci USA 2000: 97 (10): 5022–5024.
5. Sorrentino BP, Nienhuis AW. Gene therapy for hematopoietic diseases. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. W.B. Saunders: Philadelphia, 2001: 969–1003.
6. Stamatoyannopoulos G, Grosveld F. Hemoglobin switching. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 135–182.
7. Grosveld F et al. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 1987: 51 (6): 975–985.
8. Forrester WC et al. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucl Acid Res 1987: 15 (24): 10159–10177.
9. Tuan D, London IM. Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic epsilon-globin gene in K562 leukemia cells. Proc Nalt Acad Sci USA 1984: 81 (9): 2718–2722.
10. Felsenfeld G. Chromatin unfolds. Cell 1996: 86 (1): 13–19.

11. Vignali M et al. ATP-dependent chromatin-remodeling complexes. Mol Cell Biol 2000: 20 (6): 1899–1910.
12. Kingston RE, Narlikar GJ. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev 1999: 13 (18): 2339–2352.
13. Ellis J et al. A dominant chromatin-opening activity in 5′ hypersensitive site 3 of the human beta-globin locus control region. EMBO J 1996: 15 (3): 562–568.
14. Talbot D et al. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature 1989: 338 (6213): 352–355.
15. Philipsen S et al. The beta-globin dominant control region: hypersensitive site 2. EMBO J 1990: 9 (7): 2159–2167.
16. Philipsen S, Pruzina S, Grosveld F. The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. EMBO J 1993: 12 (3): 1077–1085.
17. Pruzina S et al. Hypersensitive site 4 of the human beta globin locus control region. Nucl Acid Res 1991: 19 (7): 1413–1419.
18. Talbot D et al. Detailed analysis of the site 3 region of the human beta-globin dominant control region. EMBO J 1990: 9 (7): 2169–2177.
19. Talbot D, Grosveld F. The 5′HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J 1991: 10 (6): 1391–1398.
20. Caterina JJ et al. Human beta-globin locus control region: analysis of the 5′ DNase I hypersensitive site HS 2 in transgenic mice. Proc Natl Acad Sci USA 1991: 88 (5): 1626–1630.
21. Tanimoto K et al. Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. Nature 1999: 398 (6725): 344–348.
22. Fraser P et al. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev 1993: 7 (1): 106–113.
23. Tuan DY et al. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human 'beta-like globin' genes. Proc Natl Acad Sci USA 1989: 86 (8): 2554–2558.
24. Ellis J et al. Synthetic human beta-globin 5′HS2 constructs function as locus control regions only in multicopy transgene concatamers. EMBO J 1993: 12 (1): 127–134.
25. Blobel GA et al. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. Proc Natl Acad Sci USA 1998: 95 (5): 2061–2066.
26. Boyes J et al. Perturbation of nucleosome structure by the erythroid transcription factor GATA-1. J Mol Biol 1998: 279 (3): 529–544.
27. Boyes J et al. Regulation of activity of the transcription factor GATA-1 by acetylation. Nature 1998: 396 (6711): 594–598.
28. Armstrong JA, Emerson BM. NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 *in vitro*. Mol Cell Biol 1996: 16 (10): 5634–5644.
29. Armstrong JA, Bieker JJ, Emerson BM. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF *in vitro*. Cell 1998: 95 (1): 93–104.
30. Gong QH, McDowell JC, Dean A. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene *in vivo* by 5′ hypersensitive site 2 of the beta-globin locus control region. Mol Cell Biol 1996: 16 (11): 6055–6064.
31. O'Neill D et al. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. Proc Natl Acad Sci USA 1999: 96 (2): 349–354.
32. Schubeler D et al. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev 2000: 14 (8): 940–950.
33. Gribnau J et al. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell 2000: 5 (2): 377–386.
34. Grosveld F. Activation by locus control regions? Curr Opin Genet Dev 1999: 9 (2): 152–157.
35. Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics *in vivo*. Nature 1995: 377 (6546): 209–213.
36. Dillon N et al. The effect of distance on long-range chromatin interactions. Mol Cell 1997: 1 (1): 131–139.
37. Bungert J et al. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. Genes Dev 1995: 9 (24): 3083–3096.
38. Peterson KR et al. Effect of deletion of 5′HS3 or 5′HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. Proc Natl Acad Sci USA 1996: 93 (13): 6605–6609.
39. Wood WG. The complexities of β globin gene regulation. Trends Genet 1996: 12 (6): 204–206.
40. Higgs DR. Do LCRs open chromatin domains? Cell 1998: 95 (3): 299–302.
41. Reik A et al. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. Mol Cell Biol 1998: 18 (10): 5992–6000.
42. Epner E et al. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. Mol Cell 1998: 2 (4): 447–455.
43. Bulger M et al. Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci USA 1999: 96 (9): 5129–5134.
44. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev 1999: 13 (19): 2465–2477.
45. Cone RD et al. Regulated expression of a complete human beta-globin gene encoded by a transmissible retrovirus vector. Mol Cell Biol 1987: 7 (2): 887–897.
46. Dzierzak EA, Papayannopoulou T, Mulligan RC. Lineage-specific expression of a human beta-globin gene in murine bone marrow transplant recipients reconstituted with retrovirus-transduced stem cells. Nature 1988: 331 (6151): 35–41.
47. Karlsson S et al. Expression of the human beta-globin gene following retroviral-mediated transfer into multipotential hematopoietic progenitors of mice. Proc Natl Acad Sci USA 1988: 85 (16): 6062–6066.
48. Sadelain M et al. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. Proc Natl Acad Sci USA 1995: 92 (15): 6728–6732.
49. Leboulch P et al. Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an

TAB 8

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2567–2571, March 1994
Medical Sciences

Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*

(gene therapy/gene expression/bone marrow/long terminal repeat)

PIA-MARIA CHALLITA* AND DONALD B. KOHN*†‡

Departments of *Microbiology and †Pediatrics, University of Southern California School of Medicine, Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, Los Angeles, CA 90027

Communicated by Elizabeth F. Neufeld, November 18, 1993 (received for review July 1, 1993)

ABSTRACT **We describe studies of gene transfer and expression of the human glucocerebrosidase cDNA by a Moloney murine leukemia virus (MoMuLV)-based retroviral vector in a murine gene transfer/bone marrow transplant (BMT) model. Pluripotent hematopoietic stem cells (HSCs) were assayed as the colony-forming units, spleen (CFU-S) generated after serial transplantation. Transcriptional expression from the MoMuLV long-terminal repeat (LTR) was detected at a high level in the primary (1°) CFU-S and tissues of reconstituted BMT recipients. However, we observed transcriptional inactivity of the proviral MoMuLV-LTR in >90% of the secondary (2°) CFU-S and in 100% of the tertiary (3°) CFU-S examined. We have compared the methylation status of the provirus in the 1° CFU-S, which show strong vector expression, to that of the transcriptionally inactive provirus in the 2° and 3° CFU-S by Southern blot analysis using the methylation-sensitive restriction enzyme *Sma* I. The studies demonstrated a 3- to 4-fold increase in methylation of the *Sma* I site in the proviral LTR of 2° and 3° CFU-S compared to the transcriptionally active 1° CFU-S. These observations may have important implications for future clinical applications of retroviral-mediated gene transfer into HSCs, where persistent gene expression would be needed for an enduring therapeutic effect.**

Gene therapy via bone marrow cells is a promising technique for treatment of a wide variety of human diseases, including genetic disorders, cancer, and AIDS. Effective long-term bone marrow gene therapy requires the fulfillment of two main criteria. The exogenous gene should be introduced into a high percentage of long-lived pluripotent hematopoietic stem cells (HSCs). Subsequently, the introduced gene should be persistently expressed in the mature hematopoietic progeny cells of the stem cell, thereby maintaining the effects of gene therapy for the lifetime of the individual. Although Moloney murine leukemia virus (MoMuLV)-based retroviral vectors are currently the most efficient vehicles for gene transfer into a variety of cell types including HSCs (reviewed in ref. 1), the long-term *in vivo* expression from the viral promoter/enhancer elements has been unsatisfactory. Lack of gene expression from the 5′ MoMuLV long-terminal repeat (LTR) has been observed in several systems including primary fibroblasts (2) and hematopoietic cells (3, 4). Previous studies by our laboratory, using a retroviral vector in which a normal human glucocerebrosidase (GC) cDNA is controlled by the enhancer/promoter of the 5′ MoMuLV-LTR, demonstrated a high rate of lack of expression in cells derived from HSCs (5).

Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by *de novo* methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts *in vitro* (9) and *in vivo* (2).

In this study, we investigated long-term *in vivo* expression from the MoMuLV-LTR by transduction of murine bone marrow cells with a MoMuLV-based retroviral vector and serial bone marrow transplantation (BMT) into lethally irradiated recipient mice. We document a high rate of expression failure associated with methylation of the vector LTR in the secondary (2°) and tertiary (3°) colony-forming units, spleen (CFU-S).

MATERIALS AND METHODS

Retroviral Vector. The G2 retroviral vector and its corresponding high-titer amphotropic PA317 packaging cell clone have been described (5). G2 consists of the LTR from the N2 vector flanking the human GC cDNA. The packaging cell line clone used in the experiments was negative for helper virus production assayed by testing for transfer of the amphotropic *env* gene into 3T3 fibroblasts through PCR analysis (10).

Transduction of Murine Bone Marrow Cells. Donor bone marrow cells were harvested from male C57BL/6J mice (Charles River Breeding Laboratories), prestimulated in the presence of growth factors, and cocultivated over vector producing fibroblasts according to the methods described by Weinthal *et al.* (5). The growth factors used for the prestimulation were 200 units of murine interleukin 3 (IL-3) per ml (Biosource, Camarillo, CA), 100 units of human IL-6 per ml (Amgen), 200 units of human IL-1α per ml (Immunex) and 50 ng of mast cell growth factor per ml (or c-kit ligand; Immunex).

BMT and Sample Collection. Recipient female C57BL/6J mice (8–12 weeks old) were irradiated with two split doses of 600 and 450 cGy 24 hr apart. Transduced bone marrow cells were injected into the tail vein of the irradiated mice at 1×10^6 cells per mouse for isolation of CFU-S or $2–4 \times 10^6$ cells for long-term reconstitution. Twelve days after BMT, two to four mice transplanted with 1×10^6 bone marrow cells were sacrificed. Well-defined, individual primary (1°) CFU-S were isolated and divided evenly into two portions, one for DNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MoMuLV, Moloney murine leukemia virus; LTR, long terminal repeat; HSC, hematopoietic stem cell; GC, glucocerebrosidase; CFU-S, colony forming unit, spleen; BMT, bone marrow transplantation; 1°, primary; 2°, secondary; 3°, tertiary.

‡To whom reprint requests should be addressed at: Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027.

Table 1. Expression of G2 in the mouse model of gene transfer/BMT

	Exp. 1		Exp. 2		Exp. 3		Total	
	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
1° CFU-S	9/9	5/5	12/12	6/6	8/8	8/8	29/29	19/19
1° tissues (1.5 months)								
Spleen					2/2	2/2	2/2	2/2
Thymus					2/2	2/2	2/2	2/2
Marrow					2/2	2/2	2/2	2/2
1° tissues (3 months)								
Spleen			1/1		2/2	2/2	2/2	2/2
Thymus					0/2	0/2	0/2	0/2
Marrow			1/1		2/2	2/2	3/3	2/2
2° CFU-S (1.5 months)	10/15	0/10*			3/28	0/3	13/43	0/13
2° CFU-S (3 months)			20/20	1/7	37/38	3/35	57/58	4/42
2° tissues (3 months)								
Spleen			2/2	1/2			2/2	1/2
Thymus			2/2	0/2			2/2	0/2
Marrow			2/2	1/2			2/2	1/2
3° CFU-S			7/29	0/7			7/29	0/7

*This set of 2° CFU-S was analyzed for RNA expression by reverse transcription/PCR.

and one for RNA analysis. Animals transplanted with 2–4 × 10^6 bone marrow cells were sacrificed after 1–3 months. Hematopoietic tissues were collected for nucleic acid analysis, and bone marrow cells were used directly to reconstitute a second generation of lethally irradiated female mice. Twelve days after the secondary BMT, the 2° CFU-S were isolated for DNA and RNA analysis. In one experiment, 3° BMT was performed from bone marrow of long-term-reconstituted 2° recipient animals in order to generate 3° CFU-S.

DNA and RNA Analysis. Genomic DNA was isolated by SDS/proteinase K and RNase digestion at 55°C for 3–4 hr. The digested tissues were extracted with phenol/chloroform; the DNA was precipitated in ethanol and resuspended in TE buffer. The presence of proviral GC sequences in the CFU-S and hematopoietic tissue samples was assayed by PCR using the human GC-specific oligonucleotide primers described by Weinthal *et al.* (5), followed by Southern blotting and hybridization with a ^{32}P-end-labeled internal oligonucleotide (8). Provirus DNA was also detected by Southern blot analysis after digestion of genomic DNA with the *Sst* II and *Xho* I restriction enzymes (BRL). These digestions release the 1.65-kb GC cDNA detected by hybridization with the 1.5-kb (*Sst* II/*Bam*HI) human GC cDNA probe. The probe was labeled with [^{32}P]dCTP by the random-priming method. Individual provirus integrants in the CFU-S and long-term hematopoietic tissues were detected by Southern blot analysis of genomic DNA digested with *Bam*HI, which cuts at one site in the provirus. Again, the Southern blot was hybridized with the 1.5-kb ^{32}P-labeled human GC cDNA probe.

RNA was isolated from the tissues by the acid guanidinium thiocyanate/phenol/chloroform method (11). RNA (15 μg) was electrophoresed on a 1.2% formaldehyde gel, denatured, neutralized, and transferred to a nylon membrane by capillary blotting. The filter was hybridized with the human GC cDNA probe. After a satisfactory exposure was obtained, the filter was stripped and rehybridized with the mouse β-actin DNA probe. For reverse transcription/PCR, 1 μg of RNA was reverse transcribed using the human GC-specific oligonucleotide primers, followed by PCR amplification of the cDNA as described above for the DNA samples.

Methylation Analysis. The methylation status of the proviral 5′ LTR in the CFU-S was determined by digestion of genomic DNA (15–25 μg) with *Bam*HI to reduce the size of the DNA fragments, followed by *Pvu* II digestion. The DNA was then precipitated with ethanol, redissolved in TE buffer, and divided into two equal portions, one of which was subjected to digestion with the methylation-sensitive enzyme *Sma* I. Completeness of the genomic DNA digestions was monitored by mixing a sample of the digestion mixture with either adenovirus type 2 DNA or λ DNA (BRL), which were subsequently run on a 1% gel. *Pvu* II and *Pvu* II/*Sma* I-digested DNA were electrophoresed and blotted to nylon membranes. The blots were probed with a ^{32}P-labeled fragment of the G2 vector from the *Spe* I site in the untranslated leader region to a *Pvu* II site near the 5′ end of the GC gene (see Fig. 3). Densitometric analyses were performed with the United States Biochemical SciScan 5000, measuring the relative densities of the 1.8-kb *Sma* I-resistant band and the 1.5-kb *Sma* I-sensitive band in each lane.

RESULTS

Expression of G2 *in Vivo* in Murine HSCs. Results of G2-mediated gene transfer and expression in the mouse

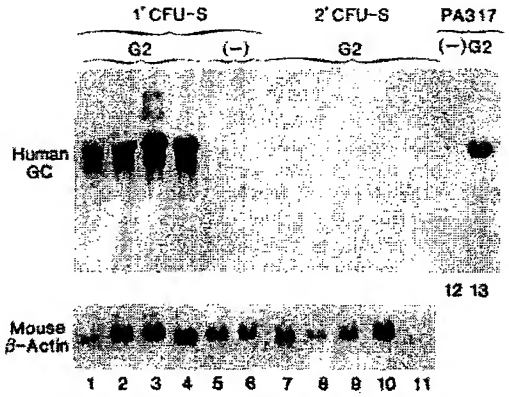

FIG. 1. Representative Northern blot analysis of the CFU-S generated in the mouse model of gene transfer/BMT (experiment 1 of Table 1). RNA from 1° CFU-S (lanes 1–6) and 2° CFU-S generated 1.5 months after primary BMT (lanes 7–11). Bone marrow cells used for BMT were transduced with G2 (lanes 1–4 and 7–11) or with the neomycin-containing control retroviral vector (lanes 5 and 6). RNAs from the fibroblast cell line PA317 (lane 12) and PA317 transduced with G2 (lane 13) were used as negative and positive controls, respectively, for GC mRNA. (*Upper*) Northern blot was probed with the human GC cDNA identifying the two proviral transcripts, the full-length 4.0 kb, and the spliced 3.5 kb. (*Lower*) Same blot was stripped and rehybridized with the mouse β-actin probe except for the control samples (lanes 12 and 13).

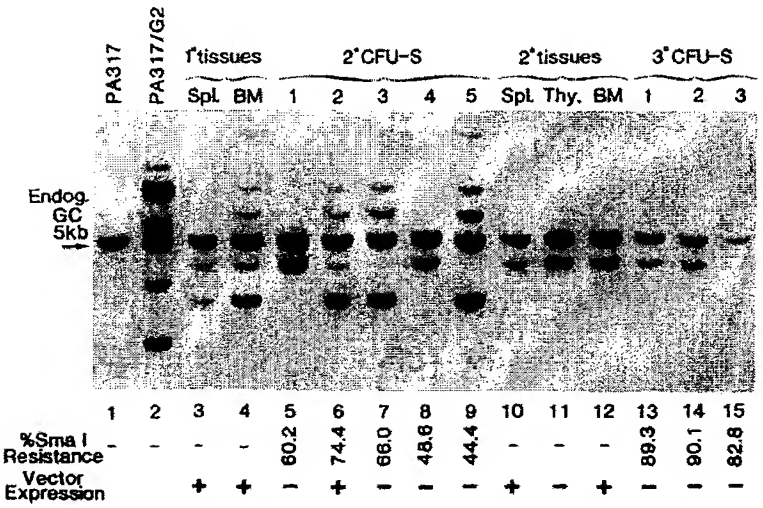

FIG. 2. Integration patterns of the G2 retroviral vector by Southern blot analysis. DNA was isolated from the tissues (experiment 2) and subjected to *Bam*HI digestion, which cuts at one site in the retroviral vector. 1° tissues refer to the spleen (lane 3) and bone marrow (lane 4) of one animal sacrificed 3 months after primary BMT. The marrow of this animal was used to reconstitute 2° recipients to generate the 2° CFU-S depicted in lanes 5–9. One reconstituted 2° recipient was sacrificed 3 months after 2° BMT. The spleen, thymus, and bone marrow (2° tissues) were analyzed (lanes 10–12). The marrow from this 2° BMT recipient was used in a 3° transplant to generate the 3° CFU-S (lanes 13–15). The extent of methylation measured for each CFU-S is indicated as the percentage of *Sma* I resistance. Vector expression was measured by Northern blot analysis and is indicated as present (+) or absent (−).

model of gene transfer/BMT are shown in Table 1. In the 1° CFU-S, a gene transfer extent of 100% was recorded in the three experiments performed. Furthermore, RNA expression was detected at a high level in all nineteen 1° CFU-S examined. An example of RNA expression from the Mo-MuLV-proviral LTR is shown in Fig. 1. The two proviral RNA transcripts in the 1° CFU-S are depicted in lanes 1–4. These data confirm previous findings from our laboratory, which demonstrated high-level expression by G2 in 1° CFU-S (5).

To confirm hematopoietic reconstitution with genetically marked cells, hematopoietic tissues were analyzed from mice 1–3 months after primary BMT. The intact human GC cDNA was detected by genomic DNA Southern blot analysis in all 10 spleen and marrow samples and in the thymus from two of four primary recipient animals. Vector RNA transcripts were detected, by Northern blot analysis, in all tissues in which the provirus DNA sequences were present (Table 1).

Analysis of the 2° CFU-S showed a high overall retroviral gene transfer frequency, with 70 of 101 (70%) containing the G2 provirus. However, RNA expression was rarely detected in these tissues. In a total of three experiments, RNA expression from the MoMuLV-LTR was detected in only 4 of 55 (7.3%) of the 2° CFU-S that contained the G2 provirus. The Northern blot in Fig. 1 shows an example of five 2° CFU-S that did not express proviral RNA transcripts (lanes 7–11). Experiment 2 was carried out further by allowing two other secondary BMT recipients to survive for 3 months after the transplantation, producing tissues stably engrafted by the serially passaged marrow (Table 1). The spleen, thymus, and marrow of the two reconstituted 2° recipients contained G2 provirus DNA by Southern blot. Vector RNAs were detected at a low level in the spleen and marrow, but not the thymus, of one of the two animals. No vector transcripts were detected in the other reconstituted 2° recipient. Tertiary BMT was also performed in experiment 2 with the bone marrow obtained from the reconstituted 2° recipients 3 months after secondary BMT. Seven of the twenty-nine 3° CFU-S analyzed contained G2 provirus DNA; these were all derived from the donor 2° animal, which had shown expression of vector RNA in its tissues. None of the seven 3° CFU-S containing proviral DNA had detectable vector RNA expression.

Gene Transfer into Pluripotent HSCs. Provirus integration patterns were followed in 1°, 2°, and 3° recipient animals to demonstrate that the 2° and 3° CFU-S, in which the LTR was inactive, are descended from true pluripotent HSCs (Fig. 2). Both the spleen and marrow of the 1° recipient, 3 months after BMT (Table 1, experiment 2), showed the same pattern, with five proviral integrants (Fig. 2, lanes 3 and 4). The five 2° CFU-S that were produced from marrow of that primary recipient showed segregation of the vector integrants seen in the primary tissues into two different patterns. The first pattern, represented in lanes 6, 7, and 9, consisted of four proviruses (although there may be slight contamination of the sample in lane 6 with DNA from a CFU-S with the integrant seen in lane 5). Interestingly, despite the same pattern of integrants among these three 2° CFU-S, there was strong vector expression seen in the one in lane 6, whereas the two foci seen in lanes 7 and 9 had no detectable vector transcripts. The second pattern of vector integrants seen in two other 2° CFU-S (lanes 5 and 8) had one provirus. The latter single integrant of ≈4 kb was also detected in the spleen, thymus, and marrow of another reconstituted secondary animal 3 months after secondary BMT (lanes 10–12). Moreover, this integrant was transferred to three CFU-S of a 3° recipient animal. Thus, the same transduced stem cell that generated the 2° CFU-S was able to reconstitute all three hematopoietic tissues (spleen, thymus, and marrow) of a secondary animal sacrificed 3 months after BMT and also to generate 3° CFU-S in the 3° transplanted animal. The provirus present in the 2° and 3° CFU-S was transcriptionally inactive (Table 1).

Methylation Analysis of the CFU-S. Genomic DNA from transcriptionally active 1° CFU-S and from inactive 2° and 3° CFU-S were compared for *Sma* I resistance according to the protocol described in *Materials and Methods* and illustrated in Fig. 3. *Sma* I is a methylation-sensitive enzyme that can cleave DNA at the CCCGGG site only if the CpG sequence is not methylated; therefore, *Sma* I resistance is used as a measurement of DNA methylation. Two examples of the resulting Southern blots are shown in Fig. 4. In the packaging cell line PA317, the G2 provirus gives a 1.8-kb *Pvu* II band (Fig. 4, lanes 1), which is reduced to a 1.5-kb band upon *Sma* I digestion (lanes 2). Complete digestion by *Sma* I shows the absence of methylation of the 5′ LTR in the fibroblast cell

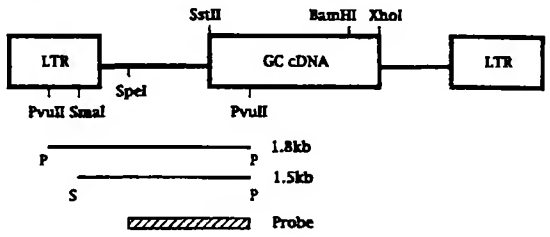

FIG. 3. Map of the G2 retroviral vector showing restriction sites and probe used for DNA methylation analysis.

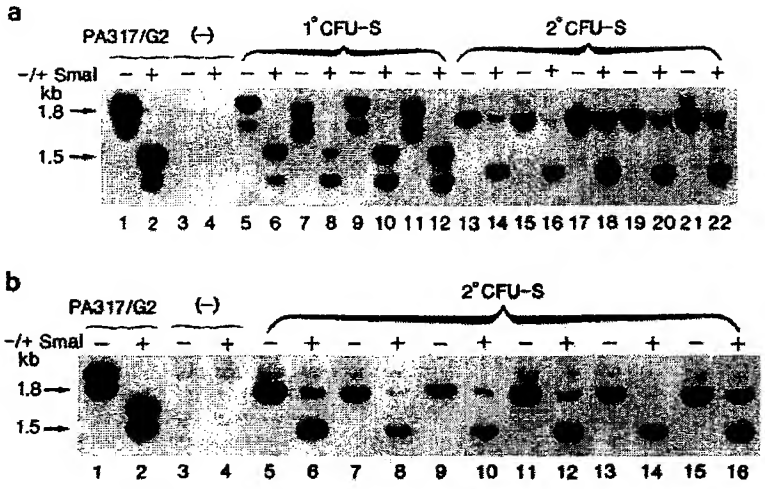

FIG. 4. Southern blot analysis of methylation of the G2 provirus in the 1° CFU-S (*a*) and 2° CFU-S (*a* and *b*). DNA was subjected to *Bam*HI and *Pvu* II with or without *Sma* I digestions. Lanes 3 and 4, labeled (−), represent DNA from normal control C57BL/6J mice. A second smaller band is seen in the PA317 cells and in the 1° CFU-S analyzed here due to the presence of a provirus with a short deletion in the 5′ untranslated region among the six copies in the packaging cell clone.

line. All *Sma* I digestions of DNA from 1° CFU-S show reduction in size of the vector-specific band to 1.5 kb, corresponding to lack of methylation of the *Sma* I site of the provirus 5′ LTR in 1° CFU-S. In contrast, the provirus in all 2° CFU-S shows some extent of *Sma* I resistance reflecting various degrees of methylation of the 5′ MoMuLV-LTR in these tissues.

To quantitate the degree of methylation at the *Sma* I site in the CFU-S, densitometric analyses were performed on the Southern blots, comparing the relative intensities of the 1.8- and 1.5-kb bands. Fig. 5 displays the percentage *Sma* I resistance of the G2 provirus in 3T3 fibroblasts, 1° CFU-S, and 2° and 3° CFU-S. In experiment 1 (Fig. 5*A*), the 1° CFU-S showed a mean of 8.9% *Sma* I resistance. In contrast, the average percentage *Sma* I resistance recorded in the 2° CFU-S generated 1.5 months after initial gene transfer/BMT was 24.6%, ranging from 9.0% to 47.6% over 10 samples analyzed. Thus, a 3-fold increase in percentage *Sma* I resistance was recorded between the 1° and 2° CFU-S of this experiment. The difference in methylation is statistically significant, with $P < 0.005$ by Student's *t* test analysis. Fig. 5*B* represents the comparison between 1° and 2° CFU-S derived 3 months after initial gene transfer/BMT in experiment 2. The percentage *Sma* I resistance was 14.9% in the 1° CFU-S and increased 3.8-fold to 56.8% in the 2° CFU-S. In the same experiment, the 3° CFU-S generated 3 months after secondary BMT showed 74.7% *Sma* I resistance, representing a 3.9-fold increase over the 19.3% *Sma* I resistance seen in other 1° CFU-S from the same experiment (Fig. 5*C*). Methylation of the provirus in both 2° and 3° CFU-S of this experiment was significantly greater than in 1° CFU-S ($P < 0.001$ by Student's *t* test). We conclude that the transcriptional inactivity observed in the 2° and 3° CFU-S is associated with methylation of the 5′ MoMuLV-provirus LTR at the *Sma* I site, 30 bp downstream of the transcription start site. Of note, the methylation status of the three 2° CFU-S with the common pattern of four vector integrants (Fig. 2, lanes 6, 7, and 9) was similar, ranging from 44% to 74%, despite the discordance for expression.

DISCUSSION

We have studied gene transfer of a MoMuLV-based retroviral vector into murine hematopoietic stem cells and expression from the MoMuLV-LTR promoter/enhancer elements in the progeny of the transduced cells. The expression from the proviral LTR was measured in the CFU-S derived from primary BMT performed after gene transfer, in the hematopoietic tissues of long-term reconstituted animals, and in the CFU-S generated after serial transplantation.

Our results demonstrate that the MoMuLV-LTR is a very efficient expression unit in 1° CFU-S. We have also detected expression from the MoMuLV-LTR in the hematopoietic tissues of transplant recipients 3 months after primary BMT.

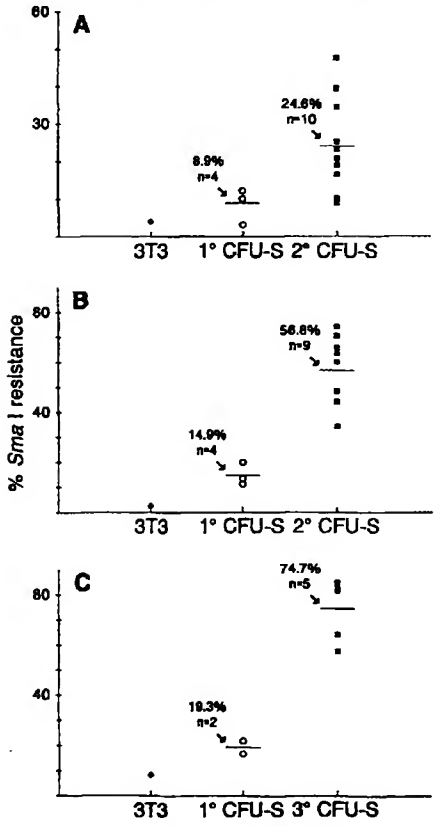

FIG. 5. Densitometric analysis to quantitate the extent of methylation of the MoMuLV-LTR in the 1°, 2°, and 3° CFU-S. The densities of the 1.8- and 1.5-kb bands after *Sma* I digestion were measured and percentage *Sma* I resistance was calculated. Each number represents the mean of three different readings. (*A*) Percentage *Sma* I resistance of samples from experiment 1. (*B* and *C*) Results from experiment 2 of the 2° CFU-S generated 3 months after secondary BMT and 3° CFU-S generated 3 months after 3° BMT, respectively. In each graph, the 1° CFU-S and either the 2° or the 3° CFU-S were analyzed from the same blot.

These results are comparable with previously published observations (4, 5, 12, 13). However, 1° CFU-S are mainly derived from committed progenitor cells restricted to the myeloid lineage (14, 15). Even the hematopoietic cells in the organs 3 months after BMT may be derived from progenitor cells capable of short-term but not long-term engraftment (16). By following the proviral integration pattern after serial transplantation, we were able to show that the cells capable of forming 2° CFU-S have the characteristics of long-lived, pluripotent HSCs. In this stringent assay of HSCs, we have observed that the MoMuLV-LTR is frequently inactive in the resultant progeny cells derived from the HSCs. We cannot, however, determine whether the absence of expression in the 2° CFU-S reflects the loss of expression by the LTR with time or whether the LTR is never active if inserted into the subclass of long-lived HSCs capable of producing 2° and 3° CFU-S.

The failure of transcription from the MoMuLV-LTR in hematopoietic tissues is in accord with prior observations (3–5, 17). However, the previous studies have mostly analyzed expression in reconstituted hematopoietic tissues of primary recipients. Moore *et al.* (17) have studied expression of human adenosine deaminase (ADA) by the MoMuLV-LTR vector expression through 2° CFU-S. Despite strong ADA expression in the primary recipients, they detected expression in only eighteen of seventy-two 2° CFU-S, although the percentage of these colonies that contained proviral DNA was not determined.

Methylation is associated with transcriptional inactivation of many genes and has specifically been seen in association with inactivity of the transduced MoMuLV-LTR in embryonic stem and embryonic carcinoma cell lines (7). Therefore, we examined the methylation status of the inactive 5′ MoMuLV-LTR in the 2° and 3° CFU-S. The presence of a high copy number of endogenous murine retroviral sequences creates a high background in the analysis of methylation across the provirus 5′ LTR. To overcome this problem, we used the restriction enzymes *Sma* I and *Pvu* II, which generate a specific provirus band detected by Southern blot analysis. Unfortunately, this assay restricted our analysis of methylation to one CpG dinucleotide contained in the *Sma* I site, 30 bp downstream from the transcription start site in the 5′ MoMuLV-proviral LTR.

Our analysis has indicated striking differences in the methylation patterns of this sequence. The *Sma* I site described is extensively methylated in the 2° CFU-S, which do not show MoMuLV proviral transcription, but is not methylated in the 1° CFU-S, which do express vector transcripts. We have observed that this same *Sma* I site is completely methylated in vector-transduced embryonic stem cell lines, which also do not show expression from the MoMuLV-LTR (data not shown). Moreover, studies done by Singer-Sam *et al.* (18) on the phosphoglycerate kinase promoter present on the inactive X chromosome have shown that methylation of a similar site, a *Hpa* II site (CCGG) at position +20, correlates with lack of transcription from the promoter. Nevertheless, the observed association between proviral methylation and expression inactivity does not show whether methylation plays a causal role in suppressing expression or is merely a secondary event after failure of expression has occurred.

Interestingly, the association between methylation and expression inactivity was not complete; at least one 2° CFU-S with extensive methylation at the *Sma* I site had a high level of vector transcripts. Other 2° CFU-S with the same proviral integrants failed to express the vector. This set of samples would indicate that neither integration site, methylation at the *Sma* I site, nor differentiation status of the transduced stem cell act to absolutely govern expression. The observed discordance suggests that commitment to expression or inactivity may stochastically be made after the HSC has differentiated to produce multiple pre-CFU-S.

The studies presented in this paper suggest that the wild-type MoMuLV-LTR may not be the ideal transcriptional unit for expression in pluripotent HSCs and their progeny cells. Successful approaches to overcoming this problem have been to use either modified LTRs with enhancer substitutions (19) or internal promoters derived from housekeeping genes (such as phosphoglycerate kinase or β-actin) instead of the wild-type MoMuLV-LTR (20, 21). Characterization of a transcriptionally active retroviral vector in HSCs may provide a better understanding of the regulation of gene expression occurring in these cells.

We would like to thank Dr. Robertson Parkman, Dr. Gay Crooks, and Dr. Punam Malik for carefully reading the manuscript; Dr. Wanda Krall for her helpful advice; and Dr. Ellen Bolotin for assistance in statistical analysis. The study was supported by National Institutes of Health Grant DK-42694 and March of Dimes Grant 1-FY92-0551. D.B.K. is a recipient of an American Cancer Society Junior Faculty Research Award (JFRA-320).

1. Mulligan, R. C. (1993) *Science* **260**, 926–932.
2. Palmer, T. D., Rosman, G. J., Osborne, W. R. & Miller, A. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1330–1334.
3. Williams, D. A., Orkin, S. H. & Mulligan, R. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2566–2570.
4. Kaleko, M., Garcia, J. V., Osborne, W. R. & Miller, A. D. (1990) *Blood* **75**, 1733–1741.
5. Weinthal, J., Nolta, J. A., Yu, X. J., Lilley, J., Uribe, L. & Kohn, D. B. (1991) *Bone Marrow Transplant.* **8**, 403–412.
6. Cedar, H. (1988) *Cell* **53**, 3–4.
7. Habers, K., Schnieke, A., Stuhlman, H., Jahne, D. & Jaenisch, R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7609–7613.
8. Jaehner, D. & Jaenisch, R. (1985) *Nature (London)* **315**, 594–597.
9. Hoeben, R. C., Migchielsen, A. A. J., van der Jagt, R. C. M., van Ormondt, H. & van der Eb, A. J. (1991) *J. Virol.* **65**, 904–912.
10. Morgan, R. A., Cornetta, K. & Anderson, W. F. (1990) *Hum. Gene Ther.* **1**, 135–149.
11. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
12. Karlsson, S. (1991) *Blood* **78**, 2481–2492.
13. Krall, W. J., Challita, P. M., Perlmutter, L. S., Skelton, D. C. & Kohn, D. B. (1994) *Blood*, in press.
14. Ploemacher, R. E. & Brons, R. H. (1989) *Exp. Hematol.* **17**, 263–266.
15. Jones, R. J., Wagner, J. E., Celano, P., Zicha, M. S. & Sharkis, S. J. (1990) *Nature (London)* **347**, 188–189.
16. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. (1986) *Cell* **45**, 917–927.
17. Moore, K. A., Fletcher, F. A., Villalon, D. K., Utter, A. E. & Belmont, J. W. (1990) *Blood* **75**, 2085–2092.
18. Singer-Sam, J., Goldstein, L., Dai, A., Gartler, S. M. & Riggs, A. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1413–1417.
19. van Beusechem, V. W., Kukler, A., Einerhand, M. P. W., Bakx, T. A., van der Eb, A. J., van Bekkum, D. W. & Valerio, D. (1990) *J. Exp. Med.* **172**, 729–736.
20. Lim, B., Apperley, J. F., Orkin, S. H. & Williams, D. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8892–8896.
21. Guild, B. C., Finer, M. H., Housman, D. E. & Mulligan, R. C. (1988) *J. Virol.* **62**, 3795–3801.